

94272

From: Chan, Christina  
Sent: Monday, May 19, 2003 9:10 AM  
To: Schnizer, Holly; STIC-Biotech/ChemLib  
Subject: RE: RUSH: seq. search in appl. no. 09/647522

Please rush. Thanks Chris

Chris Chan  
TC 1600 New Hire Training Coordinator and SPE 1644  
308-3973  
CM-1, 9B19

CRF

-----Original Message-----

From: Schnizer, Holly  
Sent: Monday, May 19, 2003 9:00 AM  
To: Chan, Christina  
Subject: RUSH: seq. search in appl. no. 09/647522

I would like to request a Rush sequence search in the above application which is After-Final.

Please search the commercial and interference databases for SEQ ID NO:5

Thank you.

Holly Schnizer  
AU 1653  
CM1-9E09  
305-3722  
mailbox: CM1-9B01

Mary Jane Ruhl  
Tech. Info. Specialist, STIC  
TC-1600  
CM-1, Room 6A-06  
Phone: 605-1155

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: 5/19/03  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

## TYPE OF SEARCH:

NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

## VENDOR/COST (where applic.)

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

DR 01-MAR-2001 (TRENBLREL. 16, Created)  
DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)  
DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)  
DE Methyl-accepting chemotaxis protein Mcpx.  
GN Mcpx.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Muschler P.F.J., Scharf B., Schmitt R.;  
RT "Methyl-accepting chemotaxis protein Mcpx of sinorhizobium meliloti.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF312879; AAC34158.1; -  
DR HSSP; P02942; 1007.  
DR InterPro; IPR000267; Asp/Glutamase.  
DR InterPro; IPR004010; Cache.  
DR InterPro; IPR004089; Cmltaxis\_transd.  
DR InterPro; IPR003660; HAMF.  
DR InterPro; IPR003661; His\_Kina.  
DR InterPro; IPR004090; Me\_chemotaxis.  
DR Pfam; PF02743; Cache; 1.  
DR Pfam; PF00672; HAMF; 1.  
DR PRINTS; PF00015; Mcpsignal; 1.  
DR PRINTS; PR00260; CHEMTNSDCCR.  
DR SMART; SM00304; HAMF; 2.  
DR SMART; SM00388; HSKA; 1.  
DR SMART; SM00283; MA; 1.  
DR PROSITE; PS00144; ASN\_GLN\_ASE\_1; UNKNOWN\_1.  
SQ SEQUENCE 788 AA; 8373 MW; 3681BC8667DE83F5 CRC64;

Query Match 4.8%; Score 112; DR 2; Length 788;  
Best Local Similarity 19.5%; Pred. No. 24;  
Matches 59; Conservative 55; Mismatches 134; Indels 54; Gaps 9;

QY 15 ITSAKRGKRDVNSLTKVETALKEAGSGNEALEGLKGIQTRPDVGAQTKIGS 74  
DB 518 LTSGVNASAEAKVAASVEVASSDAECSGEVOKAIAMOGIEQS---HEVSRIIGV 573  
QY 75 VGSALGRLNSGDATKTIISGCDIYATFTGPGVIGAGVASFVSSILSLFTGSSAKN 134  
DB 574 I-----DEIAFQTNLILNAGVBARAGAGKGFVAVAOEVRELAO--RSANAK 621  
QY 135 SVAVIDRALSKHDEA-IORHAGAKRDPFAESSAFIQVMKQGSNLTDSLSIYANVPV 193  
DB 622 EIKTLINTSAGVREGVDLVKAGALEKIAEQVQINGLIRQISSASBQA-----VGL 676  
QY 194 YKFSNFIGQLESRSQG-----AATISLSDAKRAVDFTLLYCOLVMRETLVDLAI 245  
DB 677 KEINSAYNQMDQVTOQAAMAVEETTAASMALNDEARALSLVARFQIAPQAQAASAM 736  
QY 246 LYKRGNAEHVASAVENANRYNKEADLTDFLHKLIPEQALIGAVYHPISSESKALIN 305  
DB 737 L--RGTIERKRAAAPAPENR-----PQAQPRSAAY-----SNSTORVLA 772  
QY 306 YF-307  
DB 773 KT 774

Search completed: May 19, 2003, 11:18:58  
Job time : 96 secs

```

QY 173 MKQSNLTDSLSIIANPVYKFSNFIGLESISOGAATTSLSDAKRAVDILLYCOL 232
DB 250 LKSVVOOTDKD-SEMRAMV-----FISMSKALVNO---QFAEESKCLIERA---CEI 294
QY 233 VVARETL-LVDAILYKKAHVAHAVANANRYNK--ELAADTIDELHLILPEQ-----284
DB 295 LEKKEPALPEVAEAV-----SEVAMQYESMNEFEETALSLOKTLGILEKLPQOQHSEGS 349
QY 285 --ALIG-----AVYHPISASSETSKAILNTKYEGVDPVPRPIG 320
DB 350 VSARIGWLLIFSGRVSOAVPYLESAAERLKESEFG-AKHFGVGYVYNNIG 397

RESULT 13
QYVZ08 PRELIMINARY; PRT; 617 AA.
AC 08VZ08;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative kinesin light chain protein.
GN ATG31240.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.
RA Yamada K., Banb J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,
RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
RA Yamamura Y., Yu G., Yu S., Bowser L., Carlincl P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Koeseema E., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shin P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene ATG31240 (GI:15224637).";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY063932; FAL36288.1; -
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR: 7.
DR SMART: SM00028; TPR: 8.
SQ SEQUENCE 617 AA; 68744 MW; 5894430876CFC32F CRC64;

Query Match 4.8%; Score 112.5; DB 10; Length 617;
Best Local Similarity 22.6%; Pred. No. 15;
Matches 79; Conservative 43; Mismatches 126; Indels 101; Gaps 16;

QY 25 DVNSILTKVETAL--KEAGSNPALEALEGLKEIOTRPRDVGQATKILGVSALGK 81
DB 204 DVNPVLAHVOLELANVNAAGRR--EALENKKSLEITEMFEDESKEGVANSRL-- 258
QY 82 LNSGDATKIISGCLDIAGIATTEGPGVGMIGAVASFEVSISLFTGSASKNVAVID 141
DB 259 ---ADAVAV-----LNFNALPYALKALEIHKKELOGNNAEVAOD 296
QY 142 RAL-----SKRDEAIQRH-----AAGAKRDFAESAPFQV 172
DB 297 RRLIGVYISGLEQHDKALENRLSQRVLKMGKMLLEIRAEIDANNAKVALGKYEAAID 356
QY 173 MKQSNLTDSLSIIANPVYKFSNFIGLESISOGAATTSLSDAKRAVDILLYCOL 232
DB 357 LKSVVOOTDKD-SEMRAMV-----FISMSKALVNO---QFAEESKCLIERA---CEI 401
QY 233 VVARETL-LVDAILYKKAHVAHAVANANRYNK--ELAADTIDELHLILPEQ-----284
DB 402 LEKKEPALPEVAEAV-----SEVAMQYESMNEFEETALSLOKTLGILEKLPQOQHSEGS 456
QY 285 --ALIG-----AVYHPISASSETSKAILNTKYEGVDPVPRPIG 320
DB 457 VSARIGWLLIFSGRVSOAVPYLESAAERLKESEFG-AKHFGVGYVYNNIG 504

```

```

RESULT 14
QYVZ08 PRELIMINARY; PRT; 6713 AA.
AC 09VZ08;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Bbha protein.
GN BBHA OR SA1267.
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=158879;
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE-21311952; PubMed-11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Iino T., Kanamori M.,
RA Matsunaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani U.I.,
RA Takahashi N.K., Kuhara S., Sano T., Inoue K.I., Kaito C., Sekimizu K.,
RA Hirakawa H., Kohara S., Goto S., Yabuzaki J., Kanehisa M.,
RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
RA Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL: AP003133; BAB42527.1; -
KW Complete proteome.
SQ SEQUENCE 6713 AA; 722339 MW; AF6EED226BB4888 CRC64;

```

```

Query Match 4.8%; Score 112.5; DB 16; Length 6713;
Best Local Similarity 21.7%; Pred. No. 5,5e+02;
Matches 92; Conservative 55; Mismatches 157; Indels 119; Gaps 17;

QY 16 TSAKHGRSVNSLTFVETALKEAGSN--EALR-----ALRG-----53
DB 2037 TEADRDQKQVAYNDVNTAKLLDKTAGSNQNKAVAEALQORVNTAKRALNGDARLNKXN 2096
QY 54 -LNGEIQTK-----PDRVQATKILGSVG--SALGKLN-----SGDAT 88
DB 2097 TAKQQLATWMSHLTNNAOKANLTEQIERGTIVAGVGIQANAGTLQANMQLROSTASRDKAT 2156
QY 89 KIISGCLDIAGIATTEGPGVGMIGAV-----ASFVSSILSFTG-- 129
DB 2157 KSSEDIQDANADIQNAINDAVNTNAEGTISATNNPEAMPDTINOKASQVNSAKSLNDEK 2216
QY 130 -SSAKNSVAVIDRALSKHRDEAIQRHAGAKRDFAESAPFQVYMKQS-----177
DB 2217 LAAAKQTKAKEDIGRLTDLNN--AQRTAANAEDQANLAAVNTAAKKATSLNTAMGNLK 2273
QY 178 -----NLNDS-----LSITANVPYKFSNFIGLESISOGAATTSLSDAK 220
DB 2274 HALAEKNDTRKSVYVTDADPOKQAYDTAVTQAEATNNANGSNANETVOAALNQLNOAK 2333
QY 221 RAVDFILLYCOLVYMBRETLVDLAILYKKAHVA--SAVENANRYNKELAA-DTLDPL 277
DB 2334 NDLGDNKVAQ--AKTKARALASTYSNLNNAOSTAFTSIDIQNTTYADVTYAAONANL 2390
QY 278 HKLIPQALGAVYHPISASSETSKAILNTKYEGVDPVPRPIGRRYKFTNSYWTYSIC 337
DB 2391 N-----TANGQLONGINDQNTYKQOVNFI-----DADQ--GKKQAVYNNATYNNAGIL 2435
QY 338 SEA 340
DB 2436 DKA 2438

RESULT 15
QYVZ08 PRELIMINARY; PRT; 788 AA.
AC 09F5M7;
DR 09F5M7;
SQ SEQUENCE 788 AA; 81111 MW; 5894430876CFC32F CRC64;

```

RA Wilkinson J.E., O'Byrne C., Shea T., De Loughery C., Caspi R.  
 RT "Sequence Analysis of a 101 kb Plasmid Required for Agar degradation  
 RT of a Micrococcia strain."  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF393846; AAK62869.1;  
 DR InterPro: IPR004089; Clmtaxis\_cransd.  
 DR InterPro: IPR000566; Lipocln\_cytfabp.  
 DR Pfam: PF00015; MCPsignal.1.  
 DR PROSITE: PS00213; LIPOCALIN; UNKNOWN.1.  
 KW Plasmid.  
 SQ SEQUENCE 597 AA; 65969 MW; CA09453850F918B2 CRC64;

Query Match 4.9%; Score 114.5; DB 2; Length 597;  
 Best Local Similarity 21.2%; Pred. No. 11;  
 Matches 65; Conservative 59; Mismatches 106; Indels 77; Gaps 12;

OY 16 TSARKGRSDVNSLTETALKEASGSENALEA-LEGLK-----EIQKPPRV 65  
 DB 296 TQVOVSKIDOVSSLEELKASEMRKVEITTEASVIGKNSENGHLADRINKMEHT 355  
 OY 66 GQATRIIGSGALGKLNGLGATRIISGLIVAGIATTFGGPVGIGAVASFVSILS 125  
 DB 356 AKLSKNSKMSVLSF-KSNEITRIIS----VIGSIANO-----TNLLA 394  
 OY 126 LPTGSSAKNSVAAYIDRLSKHRDEAIORHAAGAKRDEASSAFIOWKQOONLTDLS 185  
 DB 395 L--NAATEAQAQSDAGRGFVVAEE-IRKLANDSRKRETSFVNELOOQTN----- 444  
 OY 186 IIAANVYKFSNFICGLESKISGGAATTSIDAKRAVDFTLLYCOLVWRETLLVLA 245  
 DB 445 -----EANLMSMHGIRGEITTS-----AMESFQL-----I 474  
 OY 246 LYRKNAEHAASAVENANRYNKE---LAADTLDLFLKLPQALIAVYHPIASSETSK 301  
 DB 475 LQSSKDTKHAEDILSTTGQEMKLVHOVASTENTIV--VVAEQNSVGTQEIASSAELSS 532  
 OY 302 AIIATYK 308  
 DB 533 GMTNYK 539

RESULT 11  
 OY 09R99 PRELIMINARY; PRT; 306 AA.  
 AC 09R99;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
 DE FLA.  
 GN FLA.  
 OS Aeromonas punctata (Aeromonas caviae).  
 OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;  
 OC Aeromonas.  
 OX NCBI\_Taxid=648;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SCH3N;  
 RX MEDLINE=21295073; PubMed=11401962;  
 RA Raban A.A., Gryllos I., Tomas J.M., Shaw J.G.;  
 RT "Motility and the Polar Flagellum Are Required for Aeromonas caviae  
 RT Adherence to Hep-2 Cells."  
 RL Infect. Immun. 69:4257-4267(2001).  
 DR EMBL: AF198617; AAF19179.1;  
 DR InterPro: IPR001492; Flagellin.  
 DR InterPro: IPR01029; Flagellin\_C.  
 DR Pfam: PF00700; Flagellin\_C.1.  
 DR Pfam: PF00669; Flagellin\_N.1.  
 DR PRINTS: PR00207; FLAGELLIN.  
 DR PRODOM: PD000316; FLAGELLIN.  
 DR PRODOM: PD000316; FLAGELLIN\_C.1.  
 SQ SEQUENCE 306 AA; 31771 MW; 22243CB52A1D8757 CRC64;

Query Match 4.9%; Score 113.5; DB 2; Length 306;  
 Best Local Similarity 25.3%; Pred. No. 4.5;

Matches 60; Conservative 38; Mismatches 98; Indels 41; Gaps 11;  
 OY 13 LAITSARKHRKRDVNSLTETALKEASGSENALE-LEALGKGEIOTKPRVQATK 70  
 DB 72 ISLAOTAEAGMEVETGMLQRMRTLAQOOSANGSNKAKDRALKVEDQLAEINRISTATT 131  
 OY 71 IIGSV-----GSALGXKUNSG-DARKIISGCLD-----IVAGIATTFGGPVGIGAVASFVS 121  
 DB 132 FAGTKLLDGSFSGTFQVAGADANOTIGFSLAQGTGFSISGIAKAACTTIDIVSGPAGSVTT 191  
 OY 122 S--ILSFTGSSA-----KNSVAVIDRAL-----SKIRD-EAIORHAAGAKRDE 164  
 DB 192 ATGISLITFTGSGAGISISTOSKAQAVLAADAMLEVDSKRAELGAQONRLDSTRNOA 251  
 OY 165 ESSATQWKOQSNLTDLSITIANVYKFSNFICGLESKISGGAATTSIDAKR 221  
 DB 252 NISE--NVSAARSRIADAFATERANM-----TKNIILOQAASSTIAQANQ 295

RESULT 12  
 OY 09SUX2 PRELIMINARY; PRT; 510 AA.  
 AC 09SUX2;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE Putative kinesin light chain.  
 GN AT2G31240.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eucoside II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhaver G.P., Preuss D., Nielsen W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 RT thaliana."  
 RL Nature 402:761-768(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Lin X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC006593; AAD20669.1;  
 DR InterPro: IPR001440; TPR.  
 DR Pfam: PF00515; TPR.7.  
 DR SMART: SM00028; TPR.3.  
 SQ SEQUENCE 510 AA; 56494 MW; 908363937BB8F5F9 CRC64;

Query Match 4.8%; Score 112.5; DB 10; Length 510;  
 Best Local Similarity 22.6%; Pred. No. 11;  
 Matches 79; Conservative 43; Mismatches 126; Indels 101; Gaps 16;

OY 25 DVNSLTETALKEASGSENALEALGKGEIOTKPRVQATKILGVSALGK 81  
 DB 97 DVREVLAVOLELANVANGRGRE--EATEINKSLIEIKEMTFEEDSKEMGVANRSL- 151  
 OY 82 LNSGATKIISGLDVIAGIATTFGGPVGIGAVASFVSISLFTGSSAKNSVAAYID 141  
 DB 152 ---ADAVAV-----LNFEDALPYALKALEIHKRELGNNSAEVAVOD 189  
 OY 142 RAL-----SKRDEAIORH-----AAGAKRDEASSAFIOW 172  
 DB 190 RRLIGVIYSGLEQHDRALEONRLSORVLKNMGKLELIRAEIDAANMKYALGVEBAIDI 249

08T527 ID 08T527 PRELIMINARY; PRT; 877 AA.  
 AC 08T527;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE ABC transporter ABCA.1.  
 GN ABCA.1.  
 OS Dictyostellium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX4;  
 RA Anjaard C., Loomis W.F.;  
 RT "Evolution of the ABC transporter of Dictyostellium";  
 RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF491005; AAL9041.1;  
 SQ SEQUENCE 877 AA; 97375 MW; DBEC7C7ECCB343D9 CRC64;  
 Query Match 5.0%; Score 115.5; DB 5; Length 877;  
 Best Local Similarity 20.9%; Pred. No. 16;  
 Matches 77; Conservative 52; Mismatches 128; Indels 111; Gaps 17;  
 QY 7 PWFIV-----LAISAKH-----GKSDVNSLTKVETALKEAGSNEALELEGK 56  
 DB 507 PWFIVLPSTYGLSKVRHIPPYDEDEVRAITKADA-----SNRAPL-IICGLSK 559  
 QY 57 EIQT--KPDVGGATKIL-----GSVGSALGLKNSGDATK--ISGCDIYAGIATTPG 106  
 DB 560 STKLFPRKKYHAKYKLSVEKGTILGFLSNGCGKTTIGMLTGLEPAGDLYVG 619  
 QY 107 GPGVGIGAVASFSVSI-----LSLEFT-----GSSAKNSVAVIDRA-L 144  
 DB 620 HSVISINIAVRRITVSVPQHDILMAEMTAREHLOLFSEKGIPOAERESQIOKVLDOVRL 679  
 QY 145 SKHREALORHNAAGKR-----DEAESATLQWKOOSN-- 178  
 DB 680 SKISNNLSTYSGMKRRLSVAIACIGDKPIIFMDEPTTGDPSSKRHLIDLKSTKNK 739  
 QY 179 ---LTDSDL-----STIAANVPYKFSNFIGLESRIISOGAATSTLSDKRAVDFTL 227  
 DB 740 VILLISHDHEVELLADKIVINNEGVMACNGSLQKSYGEGSYNIAKSPESIPAVV 799  
 QY 228 LYQQLVY---MRETL-----VDLALYKKNAEHYASAVENANRNKELAAD- 272  
 DB 800 EFTVLSIPCKKFMKOSALQNLNFGPVTIDHQT-----ANFKOLEEITNDPNQIMRDM 854  
 QY 273 -----TLD 275  
 DB 855 SVSHSTLD 862  
 RESULT 9  
 08ZB23 ID 08ZB23 PRELIMINARY; PRT; 800 AA.  
 AC 08ZB23;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Putative Rhs accessory genetic element.  
 GN YP03606.  
 OS Yersinia pestis.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OX Yersinia.  
 OX NCBI\_TaxID=632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CO-92 / BIOVAR ORIENTALIS;  
 RA MEDLINE-21470413; PubMed-11586360;  
 RA Parthill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,  
 RA Prentice M.B., Sebatia M., James K.D., Churcher C., Mungall K.L.,  
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,

RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
 RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Kariyeh A.V.,  
 RA Leathert S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;  
 RT "Genome sequence of Yersinia pestis, the causative agent of plague";  
 RL Nature 413:523-527(2001).  
 DR EMBL: AJ14158; CAC93075.1;  
 DR InterPro: IPR001757; Artpase\_E1-E2.  
 DR PROSITE: PS00154; Artpase\_E1-E2; UNKNOWN\_1.  
 DR Hypothetical protein; Complete proteome.  
 KW SEQUENCE 800 AA; 87222 MW; 003F04A89D99298 CRC64;  
 Query Match 5.0%; Score 115; DB 16; Length 800;  
 Best Local Similarity 17.3%; Pred. No. 15;  
 Matches 101; Conservative 79; Mismatches 202; Indels 202; Gaps 24;  
 QY 12 VLAITSARHGKSDVNSLTKVETALKEA-----SGSNEAL-----EALGKGLQIKP 62  
 DB 1 MEAHDKANNNKADKTGTPTPHIVSPAAPTSGAQMTEMAGGIGDVGATGK----- 55  
 QY 63 DRVGGATKILGVSALGKNSGDAT-----KIISGC--LDIYAGIATTPGGPVG 110  
 DB 56 -RINGAAYAKTALDAKARVLDGVTPTNIVATGAPATVSGAQMTEMAGGIGDVGATG 114  
 QY 111 MGIGAVASFSVSI--LFTGSSAKNSVAVIDRALSKHDEALQIRHAGAKRDFASSAFT 170  
 DB 115 KRINAAE-----YAKTLEAKAKYLDGCVTPM--LAAGAGAGLPDSAAATIS 162  
 QY 171 QVMKQSNLTDSLSITIAANVPYKFSNFIGLESRIISOG--AATTSLSAKRAVDFTLY 229  
 DB 163 RLVKQPSGL---QFTLTASLPPQFVAVDFLSMLSSPFLVNGLASADPAVDPAVL 219  
 QY 230 ---COLVYRE-----TLVDLALTYRG 250  
 DB 220 DEDATLFTRECVLQSRITGVAVSFEQDGTGFHQRYSVIRPALMRTSLRNNAIFQOA 279  
 QY 251 NAEHYASAVENANRYNK-----ELADTLDELAKLLEPQAL----- 286  
 DB 280 SYEETITTLKENGINDFAFGFRHHPVAREFCVQOESDFDIQNLTEGFFYFEFSA 339  
 QY 287 -----IGAV-----YHP-ISASEISKALNTKYGVPDVPPIGRKRTKN 328  
 DB 340 GKNTVYADVDGSLPKGASLPYNPVNAQAQELSTITTFERSAQR--PAMVOLKQDTFKN 397  
 QY 329 STY-NTYSICS-----YANGNWFRCGSNVRNINIVSKMSQDF 367  
 DB 398 PNMMAAFSEOSGELONORPDYEHFPGRFKDAQHODEFTKRLDLRLNDANIGQASND 457  
 QY 368 YTMNSDRRLKLYTKHDG--GFW-----GTLDDEPDGOG-----HMRIFPL 407  
 DB 458 FTLPQGLFSLY--NHPRDLNHNQDLGISHGKQOMALBQASGDQGVTLNHFSEFI- 514  
 QY 408 RHGKYVSSKRWPNFMFMSSASGYIRSMENP-----GPQ 444  
 DB 515 -----HTQWTRPTPLAKPMDGPQ 533  
 RESULT 10  
 093P84 ID 093P84 PRELIMINARY; PRT; 597 AA.  
 AC 093P84;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-MAR-2002 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE MS147, putative chemotaxis protein.  
 GN Microscilla sp. PREL.  
 OS Plasmid PSD15.  
 OC Bacteria; CF8 group; Flexibacter group; Microscilla.  
 OX NCBI\_TaxID=155537;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PREL;  
 RA Zhong Z., Toudkarian A., Helinski D., Knauf V., Sykes S.,

OY 300 SKAIIYTYFGVDPVPRIGN--RRYKFTNSYMTYSICSEAYMGNYMFGCSNRYND 357  
 Db 510 ADKLLYTDY---PEEKTLITNLKRKHA-----RHICKGIARPKLV---Y 551  
 OY 358 IYVSKMSD-----GFTYMSDRKRIYTKHDQGW 387  
 Db 552 YRAKFIYVNNYKGRYNEEDTKLAKYHSLHGNMW 586

## RESULT 6

O9XOM7 PRELIMINARY: PRT: 530 AA.  
 AC O9XOM7;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Methyl-accepting chemotaxis protein.  
 GN TM1143.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
 OX NCBI\_TaxID=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MSB / DSM 3109;  
 RX MEDLINE=9287316; PubMed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,  
 Hart D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 MacDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
 Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
 RT genome sequence of *Thermotoga maritima*.  
 RL Nature 399:323-329(1999).  
 DR EMBL: AE001771; AAD36219.1;  
 DR HSSP: P02944; 1Q07.  
 DR TIGR: TM1143;  
 DR InterPro: IPR004089; Chmtaxis\_transd.  
 DR InterPro: IPR003660; HAM.  
 DR InterPro: IPR004090; Me\_chemotaxis.  
 DR InterPro: IPR000727; T-SNARE.  
 DR Pfam: PF00672; HAM; 1.  
 DR Pfam: PF00015; MCPsignal; 1.  
 DR PRINTS: PR00260; CHEMTNSDDCR.  
 DR SMART: SM00283; MA; 1.  
 DR Complete proteome.  
 SQ SEQUENCE 530 AA; 57928 MW; C04149A4F46890CE CRC64;

Query Match  
 Best Local Similarity 5.0%; Score 116; DB 16; Length 530;  
 Matches 54; Conservative 63; Mismatches 121; Indels 40; Gaps 9;

OY 15 ITSARKGRSDVNSILTKVET---ALKKSGSNEALALEBGIKGEIOTKPRVGOATRI 71  
 Db 248 ITNOLLGSKENDINSTRLESISASVOETTAOSEETISATKNIINASQAQAFADQSTQL 307  
 OY 72 LGSVSAIGKINSGDATKIISGCDIYAGIATTPGPGVGMGAVASFPSSILSTGSS 131  
 Db 308 AKKADALKKVY--EVTNIMSASADVERVSEF---OKGAETTSVETTNAL----- 356  
 OY 132 AKNSVAAYIDRALSKHRDEAIORHAAGAKRDFEASSAFIOWAKQOOSNTDPSLTIANV 191  
 Db 357 AEQNLALALNAITEAR-----AGEAGRGFAVVADEIRKLAESQOASEVNRVVEI 409  
 OY 192 -----PVYKFSNFIIGLESRIISOGAATISLDAK-----RAVDFILLYCOLVMRET 238  
 Db 410 RSIADKGVSS---ETIARVEBGKLADEADEKINSIVGAVERINEMIQNIAAIEBOT 466  
 OY 239 LLYD---LAILYRKGNAEHAASAVENANRYNKELAADT 273  
 Db 467 AADEITTAATENAKNAEETINSYKVENARLOEISAST 504

## RESULT 7

O92B35 PRELIMINARY: PRT: 1571 AA.  
 AC O92B35;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Hypothetical protein lin1716.  
 GN Lin1716.  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIP 11262 / SEROVAR 6A;  
 RX PubMed=11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 Charbit A., Cherouani F., Couve E., de Darvar A., Dehoux P.,  
 Doman E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
 Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
 Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkak G.,  
 Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,  
 Nordiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
 Rammel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
 Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;  
 RT comparative genomics of *Listeria* species.  
 RL Science 294:849-852(2001).  
 DR EMBL: AL596169; CAC96947.1;  
 DR Listlist; LIN01716;  
 DR InterPro: IPR001525; C5\_DNA\_meth.  
 DR InterPro: IPR002886; Peptidase\_M37.  
 DR Pfam: PF01551; Peptidase\_M37; 1.  
 DR PROSITE: PS00095; C5\_MTASE\_2; UNKNOWN.1.  
 DR Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 1571 AA; 167910 MW; B3BE9809D9581F6 CRC64;

Query Match  
 Best Local Similarity 5.0%; Score 116; DB 16; Length 1571;  
 Matches 84; Conservative 74; Mismatches 148; Indels 108; Gaps 18;

OY 11 IYLAITSARKGRSDVNSILTKVETALKEASGSEALALALGKGEIOTKPRVG--QA 68  
 Db 141 LVVAGSKAMFNASDINESLNVDAFKNA-----BSVAKWSKTTLDNIGLAG 190  
 OY 63 TKI-----LGSVSAIGKINSGDATKIISGCDIYAGIATTPGPGVGMGIGAVSPV--- 120  
 Db 191 TALDLAATYGDWSTSMG--LSTGEAKKMSYWDLA-----GMLASFKNID 234  
 OY 121 -----SILSLSTFG--SSAKNSVAAYIDR-----ALSKHRDEAIORHAAGAKRDFEASSA 168  
 Db 235 IDRANTLNVGVTGETELKSLIGYMTQNTLEQFALFEGAGKVAKSSTEVTKONLARE-- 292  
 OY 169 FIOVWKOOSNTDPSLTIANVPPYKFSNFIIGLESRIISOGAATISLDAKRAVDLIL 228  
 Db 293 -----KAQKILNEAKKEGKNSLAREONKLOETQARTSBC-AKVNLKDKKO--DELYR 344  
 OY 229 YCOLVARETLLVLDAILYRKGNAEHAASAVENANRY-----NKELAADTLDLFLKILPEQ 284  
 Db 345 LRYNYVMKOT-----TNAHGDFKNTSDQANNSRFSESVKLAASAGOFLLPIT--T 395  
 OY 285 ALICAVYHPIASSETSKAIIYTYFGVDPVPRIGNRRKFTNSYMTYSICSEAYM-- 342  
 Db 396 PLI-----IQAOSFTKRL-----SDIPGAVKMKKEKFPAP--EVEETVGDFFKID 439  
 OY 343 -----GNYMRGCSNRYNPRIRSKSMDGFIYTMENSDRRK 377  
 Db 440 LIPSARELAKSMGPGFTIEGVLAFKGLGIVLNTTVIPAFKAVTKFTRENDPSMK 493

## RESULT 8

Db	Accession	Length	Score	Matches	Indels	Gaps
Oy	286 LIGA-VHPIISASTSKAILN 305	305	5.18	66	99	115
Db	486 QVSAMVKHRASRLIEDAVTLH 506	506				
RESULT 4						
ID	PRELIMINARY	PRT	506 AA.			
0823M0						
AC	0823M0					
DT	01-MAR-2002 (TrEMBLrel. 20, Created)					
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)					
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)					
DE	Aeroloxin receptor protein.					
GN	SRV3395.					
OS	Salmonella typh.					
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;					
OC	Salmonella.					
OX	NCBI_Taxid=601;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=CT18;					
RX	MEDLINE=21534947; PubMed=11677608;					
RA	Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,					
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebalha M.,					
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,					
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,					
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,					
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gara P., Parry C.,					
RA	Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,					
RA	Whitehead S., Barrall B.G.;					
RT	"Complete genome sequence of a multiple drug resistant Salmonella					
RT	enterica serovar Typh CT18.";					
RL	Nature 413:848-852(2001).					
DR	EMBL; AL627278; CAD07741.1;					
DR	InterPro: IPR004089; Chmtaxis_transd.					
DR	InterPro: IPR003660; HAMP.					
DR	InterPro: IPR001610; PAC.					
DR	InterPro: IPR000014; PAS_domain.					
DR	Pfam; PF00672; HAMP. 1.					
DR	Pfam; PF00015; MCPsignal; 1.					
DR	Pfam; PF00785; PAC; 1.					
DR	SMART; SM00283; MA; 1.					
DR	SMART; SM00086; PAC; 1.					
DR	TIGRFAMs; TIGR00229; sensory_box; 1.					
DR	TIGRFAMs; TIGR00229; sensory_box; 1.					
KW	Receptor; Complete proteome.					
SO	SEQUENCE 506 AA; 55139 MW; DED2387AB045FB30 CRC64;					
Query Match 5.18; Score 118.5; DB 16; Length 506;						
Best Local Similarity 21.2; Pred. No. 4.3;						
Matches 66; Conservative 39; Mismatches 99; Indels 115; Gaps						
Oy	24 SDVNSLTKVETALKEASGNEALELEGLKEIOTKPPRVGATKILISVGSALCKLN 83	83				
Db	262 SQVSSVRNGSRLKAGNNDLNEHTRQYEVNQG-ETVYTNQMGASVYLNSETSAADKLS 320	320				
Oy	84 -----SSDAKRIISGCDIYA-----GIATTF-----GG 107	107				
Db	321 MAASAAATGAGEADMDVYKTKMDIDIAHSTORIGTITTLINDIAFOTNITLAINAAVEAARAG 380	380				
Oy	108 PYMGKIGAVNVSFVSIISLFTGSSAKNSVAVIDRALSKHDEAIFQHAAGAKRDFEASS 167	167				
Db	381 EOGKGFANVAEVEVHLAS--KSAANAADIRKLIDASATKYQSSGEVYHAAGRITD----- 433	433				
Oy	168 AFIOVMKQOSNLWTDSDLSITAAVNPVYKFSNFIGOLESRISQGA--AFITSLDPAKRAVDF 225	225				
Db	434 ---DIVAQOVAVT-----LLIARIISQSQFOETDGLISLITRAVD- 468	468				
Oy	226 ILIVCOLVYMKETLLVDALILYRKGNAEHVASAVENMNNRVAKELAAOTLDFIHLIPEOA 285	285				
Db	469 -----ELNRITQKNA-----LVESA 485	485				

QY	286	LIGA-VYHPISASETSKAIIN	305
Db	486	QVSAWVKHRA\$RL\$EDATVTLH	506
RESULT 5			
ID	062187	PRELIMINARY;	833 AA.
AC	062187		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Transcription factor.		
GN	TFP1 OR TTF-1.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxId=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=EHRLICH ASCITES.		
RC	MEDLINE=95237206; PubMed=7720715;		
RA	Evers R., Smid A., Rudloff U., Lottspeich F., Gummt I.,		
RT	"different domains of the murine RNA polymerase I-specific terminatio		
RT	factor TTFP1 serve distinct functions in transcription termination."		
RT	EMBO J. 14:1248-1256(1995).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=EHRLICH ASCITES.		
RC	MEDLINE=95320168; PubMed=7597036;		
RA	Evers R., Gummt I.;		
RT	"Molecular evolution of mammalian ribosomal gene terminator sequences		
RT	and the transcription termination factor TTF-1,"		
RL	Proc. Natl. Acad. Sci. U.S.A. 92:5827-5831(1995).		
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).		
CC	-1- SIMILARITY: CONTAINS 2 MYB-LIKE DOMAINS.		
DR	EMBL; X83974; CA58808.1; -		
DR	MED; MG1:105044; Tef1.		
DR	InterPro: IPR001005; MYB DNA-binding.		
DR	Pfam; PF00248; myb-DNA-binding; 1.		
DR	SMART; SM00395; SANT; 2.		
DR	PROSITE; PS50090; MYB_3; 1.		
DR	DNA-binding; Nuclear protein.		
SO	SEQUENCE 833 AA; 94534 MW; DEB9C8DC6D3ECC6E CRC64;		
Query Match	5.08; Score 117; DB 11; Length 833;		
Best Local Similarity	22.08; Pred. No. 12; Mismatches 127; Indels 174; Gaps		
Matches 100; Conservative 54;			
4 KHL-----PWLEIVLAITSAKHGRSDVNSLTTRKTEALKEASG-SNEAALEALEGLK---	55		
235 KHLQNVAPW-----DVVOGSPPESTISLPSPPLSSSEDELEKSTAAVFCRSLKKNV	286		
56 --GEIQTRPDVGOATKIIIGSVSALGKINSGDATKIIISGLDIYAGIATTFGPGYMG	112		
287 FRSGLELRIPD-----SLDSETERLD-----STHGAGVAG	321		
113 -----IGAVASPYSSILSLFTGSSAKNSVAAYDR-----	142		
322 ECSESTRESHISIKKSKKKHKSVALATSSDSASVATDSKAN---ALVDSESGGAVARED	378		
143 -----ALSKHDEAIOR-----HAAGAKRPAFASSAFIYQMKOOSMLTSDLS	185		
379 VDHRAPEAEACSTEKHR-EAMKRLPETHHEESNNSASNAARIHSDRRRSDSDVD	437		
438 LGSAA--VROLRREFIPIOER--AAAT-----TRRYRDLG-469			
186 IIAANPYKFSNFIGOLESISOGAATSLSDAKRAVDITLLYCOLVMREFLYDLAI	245		
246 LYRGNMEHVA-----SAVENANRVKKELAADFLDFLKLIPQALIGAVYHPISASET	299		
470 LFEKKAGGVAIRGKRSAAK--NKQIKNVODL-----SLTIGES	509		

```

OY 138 AYDR--ALSKHDEAIOHHAAGKRDFAESSAFIOVMKQSNLTDSDLS--TIANVY 193
DB 414 AVGEAGVALGKHIGSSIRKGAADYTNLNESS-----YPELDEDIADQADATPEEN 464
OY 194 YKFSNFIQGLRESRISGAATTSISD-----AKRAVDILLYC-----OLVAKRETLVDL 243
DB 465 WADNPNSAVSRRLQAFPMASDAEIAKRAKVAATNSDQSRFTNORQDQNNKKT----- 519
OY 244 ALYKRGNAEHVASAVENANRVNKEADTLDFLKLHLEPQALICAVHPISASETSK-- 301
DB 520 ATPY--GNARDLVYNATNA--FOKYEGBDKDTFMSQLEPN-----MSAEKEKQW 566
OY 302 -----ALINY-----TKYFGVP--DVPRPIGNRRYKFTSNYNTSICEAATMGNNM 346
DB 567 NDHLNAKVOGFERNHAEQATKAGAMPVADAKQKQNN--LFDKSYNKKDPAISQALG--- 621
OY 347 FRGGSNVANPNIRVSKMDGF--TYMENSDR-----RKLYTTKHQDQWG 388
DB 622 -----KYGKAVEGVKGETLESQHONMGVGLVGAFTAFKKGTYADHKAGF- 668
OY 389 WGLDEPDGQDHMRFIPLRHGKYVSSKRWPMNFMVMESSASGR 433
DB 669 ---WKQFPAD-----MSQOEKAAMKHLIDQKVGQF 696

```

## RESULT 2

```

OY 082G07 PRELIMINARY; PRT; 595 AA.
AC 082G07:
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE Putative methyl-accepting chemotaxis protein.
GN Yersinia pestis.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OY NCBI_Taxid=632;
RN 11)
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Baker S., Basham D., Bentley S.D., Brooks K.D., Churcher C., Mungall K.L.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Hollroyd S., Jagsels K., Kariyasev A.V.,
RA Leather S., Moutle S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DB EMBL: AJ414147; CAC90064.1; -
DR InterPro: IPR004010; Cache.
DR InterPro: IPR004089; Chmtaxis_transd.
DR InterPro: IPR003660; HAMF.
DR Pfam: PF02743; Cache; 1.
DR Pfam: PF00672; HAMF; 1.
DR Pfam: PF00015; MCPsignal; 1.
DR SMART: SM00304; HAMF; 1.
DR SMART: SM00283; MA; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 595 AA; 61787 MW; 6D8962C8F7B2E5A CRC64;

```

Query Match 5.1%; Score 119.5; DB 16; Length 595;  
 Best Local Similarity 23.8%; Pred. No. 4.7; Indels 41; Gaps 11;  
 Matches 57; Conservative 39; Mismatches 103;

```

OY 21 GKRDVNSLTKVETALKEASG-----SNEALEALEGLK--GEIOTKPDVQ 67
DB 371 GKRESAASLQQTSALEQISATVAQASASARQANNVAFSSSEDAASGDVTR----- 424
OY 68 ATRILGSVGSALGKINGDTRKIISS-----CIDIYAGIATTTGGGVGAGIVASVFS 121

```

```

DB 425 VITMESIERKASGKI--GDITSVIDGIAFOQINILANVAEAREAGEGRGAVAGEV- 481
OY 122 SLTSLFTGSSAKNSVAVIDRALSKHREAIQ--RHAAGAKRDPAES--SAFIOVMKQSNL 179
DB 482 RILAOBSAQAAK--EIKALIESVSSVAGSSQVROASNMATDIVASVDVITIMSEITNA 540
OY 180 TDSDSLITIAANVPYKFSNFIQGLRESRISGAATTSISDAKRAVDPIILYCOLVMMRETL 239
DB 541 ADEQMR-----GHEINSVAQDLDTMVOQNALVQESTTASAA---LQAOADLTDIV 590

```

## RESULT 3

```

OY 082LX8 PRELIMINARY; PRT; 506 AA.
AC 082LX8:
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE Aerotaxis sensor receptor, senses cellular redox state or proton
DE motive force.
GN AER OR STM3217.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OY NCBI_Taxid=602;
RN 11)
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGGC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
DB EMBL: AE008848; AL22090.1; -
DR InterPro: IPR004089; Chmtaxis_transd.
DR InterPro: IPR003660; HAMF.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS_domain.
DR Pfam: PF00672; HAMF; 1.
DR Pfam: PF00015; MCPsignal; 1.
DR Pfam: PF00785; PAC; 1.
DR SMART: SM00283; MA; 1.
DR SMART: SM00086; PAC; 1.
DR TIGRFAMS: TIGR00229; sensory_box; 1.
DR TIGRFAMS: TIGR00229; sensory_box; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 506 AA; 55063 MW; 70C37C6E7C1BD150 CRC64;

```

Query Match 5.1%; Score 118.5; DB 16; Length 506;  
 Best Local Similarity 21.2%; Pred. No. 4.3; Indels 115; Gaps 11;  
 Matches 68; Conservative 39; Mismatches 99;

```

OY 24 SDVNSLTKVETALKEASGNEALEALEGLKGEIOTKPDVQATKILGSVGSALGRLN 83
DB 262 SQVSSVRNGSRILAKGNNDLEHTRQYVENVO--ETVTTNMASVSKINSETASADKLS 320
OY 84 -----SGDATRIISGCLDIYA-----GIATTF-----GG 107
DB 321 MAASSAATOGGEANDVYIKTMDIDIAHSTQRGTTITLINDIAFOQTNILALNAVAEARAAG 380
OY 108 PYVGIGIAVASFVSIISLTGSSAKNSVAVIDRALSKHREAIQ--RHAAGAKRDPAES 167
DB 381 EOGKGFVAVVAGEVRLHAS--RSANAAANDIRKLIDASATKYQSGSEGVHAAGRTMD----- 433
OY 168 AFTQVMKQSNLTDSDLSITIAANVPYKFSNFIQGLRESRISGA--ATISLSDAKRAVD 225
DB 434 ---DIVAQVQNTA-----LTIARISQSTOEOTDLSILTRAVD- 468
OY 226 ILLXCOLVMMRETLVDLAILYKRGNAEHVASAVENANRVNKEADTLDFLKLHLEPQA 285

```



GenCore version 5.1.4-P5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2003, 11:14:48 ; Search time 89 Seconds  
(without alignments)  
1041.811 Million cell updates/sec

Title: US-09-647-522-5

Perfect score: 2322  
Sequence: 1 MILKHPFLVIAITSAKH.....SGYRSWENPQGHMSIT 450

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL.21.\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.todent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*  
15: sp.virus:\*  
16: sp.bacteriophage:\*  
17: sp.archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	137	5.9	1222	2	Q9X349 bacillus an
2	119.5	5.1	595	16	Q82GQ7 yersinia pe
3	118.5	5.1	506	16	Q82LX8 salmonella
4	118.5	5.1	506	16	Q823M0 salmonella
5	117	5.0	833	11	Q62187 mus musculu
6	116	5.0	530	16	Q9X0M7 thermotoga
7	116	5.0	1571	16	Q92B35 listeria in
8	115.5	5.0	877	5	Q8T5Z7 dictyosteli
9	115	5.0	800	16	Q82B23 yersinia pe
10	114.5	4.9	597	2	Q93P84 microscilla
11	113.5	4.9	306	2	Q9R9R9 aeromonas p
12	112.5	4.8	510	10	Q9SUX2 arabidopsis
13	112.5	4.8	617	10	Q8VZ08 arabidopsis
14	112.5	4.8	6713	16	Q99U54 staphylococ
15	112	4.8	788	2	Q9F5M7 rhizobium m
16	112	4.8	788	16	Q92SH9 rhizobium m

17	112	4.8	2026	2	Q9JPJ4 neisseria m
18	111.5	4.8	655	5	Q94479 dictyosteli
19	111.5	4.8	656	16	Q9WYN0 thermotoga
20	111	4.8	865	2	Q9XDU1 clostridium
21	111	4.8	1215	16	Q8XJ10 leptomonas
22	111	4.8	1657	5	Q9BJ59 thermotoga
23	110.5	4.8	656	16	Q9WYR0 thermotoga
24	110.5	4.8	661	16	Q9X020 haifa alive
25	110.5	4.8	1061	2	Q8RNY3 halobacteri
26	110	4.7	810	17	P82857 anabaena sp
27	110	4.7	2297	16	Q8YLO8 haemophilus
28	109.5	4.7	874	2	Q06277 staphylococ
29	109.5	4.7	6713	16	Q931R6 staphylococ
30	109	4.7	541	16	Q9HUP8 pseudomonas
31	109	4.7	612	16	Q9RXX1 deinococcus
32	108.5	4.7	456	2	Q93UJ9 synecococc
33	108.5	4.7	659	16	Q9KVP3 vibrio chol
34	108	4.7	492	16	Q8UHY4 agrobacteri
35	107.5	4.6	387	2	Q86266 agrobacteri
36	107.5	4.6	2155	16	Q8UPP9 mus musculu
37	107	4.6	482	11	Q923V9 agrobacteri
38	106.5	4.6	692	16	Q8U1B0 borrelia bu
39	106.5	4.6	1098	16	Q507J3 mycoplasma
40	106.5	4.6	3183	2	Q8RR93 neisseria m
41	106	4.6	1975	16	Q9K0S7 neisseria m
42	105.5	4.5	387	2	Q86265 pseudomonas
43	105.5	4.5	667	16	Q8XRM8 ralsitonia s
44	105.5	4.5	2504	2	Q85160 photorhabdu
45	105	4.5	384	2	Q05492 bacillus ce

## ALIGNMENTS

## RESULT 1

ID Q9X349 PRELIMINARY: PRT: 1222 AA.

AC Q9X349: 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE PX01-79.  
OS Bacillus anthracis.  
OC Bacillus anthracis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1392;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-STERNE:  
RA MEDLINE-99445483; PubMed-10515943;  
RA Okinaka R.T., Cloud K., Hampton O., Hoffmaster A.R., Hill K.K.,  
RA Keim P., Koehler T.M., Lamke G., Kumano S., Mahillon J., Manter D.,  
RA Martinez Y., Rieke D., Svensson R., Jackson P.J.;  
RT "Sequence and organization of px01, the large Bacillus anthracis.  
RT plasmid harboring the anthrax toxin genes."  
RL J. Bacteriol. 181:6509-6515(1999).  
DR EMBL: AF065404; AAD32383.1; --  
KW Plasmid.  
SQ SEQUENCE 1222 AA; 130407 MW; A7C117874D097E5B CRC64;

Query Match 5.9%; Score 137; DB 2; Length 1222;  
Best Local Similarity 23.0%; Pred. No. 0.83; Mismatches 173; Indels 130; Gaps 23;  
Matches 107; Conservative 55;

OY 21 GKRSDV-NSLTFKVFETALKEAGSNEALEGLKEGIOTFDRVGATKIL--GSVGS 77  
DB 310 GRRDVRNRSKVRPDDGDKER--TKEDADNPDAALGANIGTIGTSRRARHLKAGVGS 367  
OY 78 ALGKLNSGDATKTIISGCLDIIVAGIATTFGGVPMGIGAVASFYSSILSLFTGSSANNSVA 137  
DB 368 ALGKATGTG-----LAGMAA--GAGLPGAMVAVSGVSGALGAPGAAGRSVA 413

RT "characterization of the type a flagellin gene from *Pseudomonas*  
RT *aeruginosa* PAK.";

RL J. Bacteriol. 172:7188-7199(1990).

RM [2]

RP SEQUENCE OF 1-389 FROM N.A.

RC STRAIN-ATCC 27853;

CC Tunngpradabkul S., Senapin S., Panyim S.;

CC "PCR-based method for isolation of the flagellin genes from

CC *Pseudomonas* species.";

CC J. Gen. Appl. Microbiol. 44:239-242(1998).

CC [3]

CC SEQUENCE OF 46-385 FROM N.A., AND GLYCOSYLATION.

CC STRAIN-5939, and 170018;

CC PubMed-9620973;

CC Birmer C.D., Montie T.C.;

CC "Cloning and comparison of *fliC* genes and identification of

CC glycosylation in the flagellin of *Pseudomonas aeruginosa* a-type

CC strains.";

CC J. Bacteriol. 180:3209-3217(1998).

CC [4]

CC PHOSPHORYLATION.

CC STRAIN-PAK, 2993, 5940, and 170018;

CC PubMed-7682214;

CC Kelly-Wintenberg K., South S.L., Montie T.C.;

CC "Tyrosine phosphate in a- and b-type flagellins of *Pseudomonas*

CC *aeruginosa*.";

CC J. Bacteriol. 175:2458-2461(1993).

CC -1 FUNCTION: Flagellin is the subunit protein which polymerizes to

CC form the filaments of bacterial flagella.

CC -1 PPM: Phosphorylated on tyrosine residue(s).

CC -1 PPM: Flagellin from strain 5939 but not from strain 170018 is

CC glycosylated.

CC -1 SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>

CC or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).

[illegible]

Query Match	4.3%;	Score 101;	DB 1;	Length 393;
Best Local Similarity	22.7%;	Pred. No. 7.1;		
Matches	73;	Conservative	54;	Mismatches 130;
				Indels 64;
				Gaps 14;

[illegible]

Db 130 TFGGRKLLDSEFGVAFQVGSAA-ANELIISGIDEMSAEISLNGTYTFFADGGAGVATATAG 188

Qy 108 PYVGIGIGAVAFSVSSILSTFTGSSAKNSVAAVIDRALSKHREDAIORHAAQARD----- 162

Db 189 TVDLAIG-----ITGGSANV---VKYDMKGNFTAEQAAPKAIK-AVNDANVGI 232

Qy 163 --FAESSAFIOVMKQOSNLSDLSITAAVNPVYKFSNFIGOLE-----SRISQGA 211

Db 233 GAFSGDGTITSVYSGAKDGSGAITSNVSQVIADTGSTGTAGAVTPSPATAFAKTIDTV 292

Qy 212 ATTSLSDDAKRAAVDFLLIYCOLVMRETLVDLAILLYK-----GMEHVASAVENA-NRV 265

Db 293 AKIDISTAKGQASAVLVYDEAIKQIDQORADLGAVQNRFNTINNKNIGDENVSAAAGRI 352

Qy 266 -NKEIADTDLELHKRIPEQA 285

Db 353 EDTDFRAETANTIKNOVLOQA 373.

Search completed: May 19, 2003, 11:16:53  
Job time.: 32 secs

DB 714 EYLGRKESIPERKVAACVHAGGSGKSHAIQALREIGSGDITVLPTELRLMWSK 773  
 QY 100 GIATP-----FGG-----PVGMGIGAVASFSVI--LSLFTGSSAK 133  
 DB 774 KVNTEPEYMERKTEKALIGGTGIVIFDDYSKILPRG--IALVSFKIKILITITGSSRQ 832  
 QY 134 NSAAVAYDRALSKHREDAIQ-----RHAAGA-----KRDAESSAFTQVKKOOSNLTDSD 183  
 DB 833 SVYHETSDDASIRHILGAPATEVFAKCYRYLYNATFRHKKDLAN--MLGVYSERTGTETIS 889  
 QY 184 LSI-----IAANP-----YKFSNFTGOLSR--SOGAA 212  
 DB 890 MSEEFLGCVPTLVPSDEKRLVMGTGRNDFTTYAGCGGLKPVQVYLNDHTOVCSANVM 949  
 QY 213 TTSLDAKRAVDFI-----LLYCOLVARETLVLVDLALYRKGAHEV 255  
 DB 950 YTLASRATDRHINTNSANSAPWEKLDSPYLTPLTSVREQALRE--YEPVEAPEI 1005  
 QY 256 AS-----AVENANRYNKELAADTID-----FLHKLIP 282  
 DB 1006 REPEPOTHMVCEVNEESLEEEYKELEKEDREIHEAHGSHNCVQETEDTVQLFSSHQAK 1065  
 QY 283 EQALIGAVYHPISASETSKA-----ILNTYKYGVDVPPPIGNRRKFT 327  
 DB 1066 DETLLAATTDARKLTSIQESNFRFETLSKRDIGVLEPLNTQKANGLPKEPPI-----FS 1118  
 QY 328 NSYMTNTYISCEAVYMGNYMFGCCSVNPNIRVSKMSDGFYTMENSDRRKLYITKHDGM 387  
 DB 1119 OEWEA---CAHEVQSKYLSKSCNKLINGVRS-----PDDEKINWFLKSO-- 1164  
 QY 388 GWCITDEDGDOGHMFIP--LRHGKYM-----SKRWPMFMYMESSASGYR 435  
 DB 1165 -WVTKYKELG-----LEKIKPGQITIAFYQOTVLMFGMARYMFMFOAQPKREVEI 1215  
 QY 436 SWENNGPGQGHNSIT 450  
 DB 1216 NCETTPEDMSANALS 1230

RESULT 14  
 ID PYRB\_ARCFU STANDARD: PRT: 299 AA.  
 AC O30130:  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Aspartate carbamoyltransferase (EC 2.1.3.2) (Aspartate transcarbamylase) (ATCase).  
 GN PYRB OR AF0106.  
 OS Archaeoglobus fulgidus.  
 OC Archaea: Euryarchaeota: Archaeoglobi: Archaeoglobales:  
 OC Archaeoglobaceae: Archaeoglobus.  
 OX NCBI\_TaxID=2234;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=96049343; PubMed=9389475;  
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Kechum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
 RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.:  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon Archaeoglobus fulgidus.";  
 RL Nature 390:364-370(1997).  
 CC -I- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-aspartate -> phosphate

CC + N-carbamoyl-L-aspartate.  
 CC -I- PATHWAY: Pyrimidine biosynthesis; second step.  
 CC -I- SUBUNIT: Heterooligomer of catalytic and regulatory chains.  
 CC -I- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE001099; AAB91125.1; -  
 CC HSP: P00479; 3CSU.  
 CC TIGR: AF0106; -  
 CC InterPro: IPR002029; Asp/Otn\_Cotranf.  
 CC InterPro: IPR002082; Asp\_CarbMltransf.  
 CC Pfam: PF00185; OTCace; 1.  
 CC Pfam: PF02729; OTCace\_N; 1.  
 CC PRINTS: PR00100; AOTCASE.  
 CC TIGRFAMs: TIGR00670; asp\_carb.tr; 1.  
 CC PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.  
 CC Pyrimidine biosynthesis; Transferase; Complete proteome.  
 CC SEQUENCE 299 AA; 33633 MW; 70F2FC2EBACDD720 CRC64;  
 KW

Query Match 4.4%; Score 101.5; DB 1; Length 299;  
 Best Local Similarity 22.6%; Pred. No. 4.6;  
 Matches 75; Conservative 38; Mismatches 122; Indels 97; Gaps 16;

QY 23 RSDVNSILTKVETALKENAGSNEALEALG--LKGEIOTKPRVGOATILSGSALGK 81  
 DB 13 REDINVLKAKEFEEDVANG--EKKRILEGKILGLNLFEP-----STRHSFETAMR 65  
 QY 82 LNSGDAATKIIISGCLDIVAGIATTFGGVGMGIGAVASFSYLSLFTGSSANNSVAAYID 141  
 DB 66 L-----GADV-----INMTAQEASSI-----ANGETLADTI 91  
 QY 142 RALSKRDEAIIORHAAGAKRDEASAFIQVK-----QSNLUDSDLSITIANVPYK 195  
 DB 92 RYVSGICDAIVTRHPEGAPRAENSVPVINAGDAGQHPOTYLLD-----YT 142  
 QY 196 FSNFTIGOLESRISOGAATTSLSDAK--RAVDPIILYCOLVYAR-----ETLV 241  
 DB 143 IKKEGRLD-----GITLIMGDLKYSRTIHSILKALMLFDMRIVLISPEALALPDITL 197  
 QY 242 DLAIILYKRGMAHVASVNE--NANVYNKELADTIDFLHLK-----IPQALIGA----- 289  
 DB 198 DVSAEIRRAAREVISEIDVLYVTRIQKRFPEDEEY--RKVSGSYRITAEITLSAKDSMI 256  
 QY 290 VYHPI-----SASETSKALINTYKFGVP 313  
 DB 257 VMHPLPRVDEIHPSVSTIKHARYFOOSFYGVF 288

RESULT 15  
 ID FLAA\_PSEAE STANDARD: PRT: 393 AA.  
 AC P21184; O30388; O33928; O68382;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE A-type flagellin.  
 GN FLIC OR FLAA.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria: Proteobacteria: gamma subdivision: Pseudomonadaceae;  
 CC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N. A., AND SEQUENCE OF 1-5.  
 RC STRAIN=PAK;  
 RX MEDLINE=91072275; PubMed=2123866;  
 RA Totten P.A., Lory S.;

```

Db 410 LIDASADKVGSGSOQVHAAGRTME-----DIVAQYKNVT----- 443
OY 199 FIGUESISOGA--ATSLSDAKRAVDPIILYCO--LVARETLVLDAIILYKGNABH 254
Db 444 ---OLIAQISHSTLEADGLSITRAVDLMLITOKNAIVEESAOVSAMV-----KH 493
OY 255 VASAVENA 262
Db 494 RASRLEDA 501

RESULT 12
RRPO_PVXCP STANDARD: PRT: 1456 AA.
ID RRPO_PVXCP
AC P22591;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE RNA replication protein (165 kDa protein) (ORF 1) [Contains: RNA-
directed RNA polymerase (EC 2.7.7.48); Probable helicase].
OS Potato virus X (strain CP) (PVX).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.
OX NCBI_TaxID=12184;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90364772; PubMed=2392880;
RA Orman B.E., Celnik R.M., Mandel A.M., Torres H.N., Mentaberry A.N.;
RT "Complete cDNA sequence of a South American isolate of potato virus
X."
RL Virus Res. 16:293-306(1990).
CC -1- FUNCTION: RNA-replication. The central part of this protein
CC possibly functions as an ATP-binding helicase.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA)(N).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M31541; AAA47178.1;
DR EMBL: X55802; CA339324.1;
DR PIR: S14005; S14005.
DR InterPro: IPR001788; RNA_dep_RNAPol2.
DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
KW ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
KW Transferase.
FT NP_BIND 735 742 POTENTIAL.
SQ SEQUENCE 1456 AA; 165301 MW; 489BA57EA070BD2E CRC64;

Query Match 4.4%: Score 103; DB 1; Length 1456;
Best Local Similarity 18.7%: Pred. No. 28;
Matches 104; Conservative 70; Mismatches 187; Indels 194; Gaps 26;

OY 52 EGLKGELQTRDVRGQATKILGVS-----ALGKLNSG-DATKIISG---CLDIYA 99
Db 714 EYLNGKISLPERKVAACVHAGSGSGSHAIKALREIGSGSDITVLPLNLEALDMSK 773
OY 100 GIATT-----FGG-----PYMGIGAVASVSSI-ISLFTGSSAK 133
Db 774 KVPNTEPYMFTYERKALIGTGSIIVFDYSKLPGY-IEALVSPSTIKILITLGDGSRQ 832
OY 134 NSVAVIDRAISKHDEAIO-----RHAAGA-----KPDFESSAFIQMGOOSLTDSD 183
Db 833 SVYHETSDASIRHLGPAEVPFAKCYRYLNATHRNKRDLAN--MLGVSERTGTTEIS 889
OY 184 LSI-----IAANVP-----VYKFSNFIGOLESRI-----SOGAA 212

```

```

Db 890 MSSEFLGCVPLTVPSEDEKRLLYMGTRNDFTYAGCOGLTKPKYQIVLDHNTQCSANVM 949
OY 213 TTSUSDARAVDFI-----LYICQLYMNETLLVDLAIIYKGNABH 255
Db 950 YSALSRAIRDRHFTINTSANSAPFEKLDSTPYLTKFTLSVREHALKE---YEPAAEAPI 1005
OY 256 AS-----AVENANRVKLEADPLD-----FLHKLIP 282
Db 1006 KEPEPOTHCYENESVLEETKEELEKFEDEIHSDAHGNSCVQTEDTTQILSHQBAK 1065
OY 283 EQALIGAVYHPIISASETSA-----ILNRYTFGVDPVPRIGNRRRYKT 327
Db 1066 DETLLMATIDARLKTNSQESNREFELSKRDIGVLFLLYKQAMGLPKPEIP-----FS 1118
OY 328 NSYNTYSIGCEAYMGVNGMFGCCSNVRNPINRVSKMSGFTTMSDRKLYITKHQDGW 387
Db 1119 QEVWEA---CAHEVOYKLSKSKCNLNGTVROS-----PDFENKIMVFLKRSO-- 1164
OY 388 GMGTLDEDPGQGHMRFIP-LRHGKVMY-----SSKRPNMFMYESSASGYR 435
Db 1165 -WTKVEKLG-----LPKIKPGOTIAFYQOYVLMFGTMARKYRMRFOAPKREYFI- 1215
OY 436 SMENNPGPQGHMSIT 450
Db 1216 NCETTPEDMSAMALS 1230

RESULT 13
RRPO_PVXHB STANDARD: PRT: 1456 AA.
ID RRPO_PVXHB
AC Q07630;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE RNA replication protein (165 kDa protein) (ORF 1) [Contains: RNA-
directed RNA polymerase (EC 2.7.7.48); Probable helicase].
OS Potato virus X (strain HB) (PVX).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.
OX NCBI_TaxID=73488;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94015010; PubMed=8409947;
RA Querol M., van der Vliugt R., Goldbach R., Salazar L.F.;
RT "RNA sequence of potato virus X strain HB."
RL J. Gen. Virol. 74:2251-2255(1993).
CC -1- FUNCTION: RNA-replication. The central part of this protein
CC possibly functions as an ATP-binding helicase.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA)(N).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X72214; CA51012.1;
DR InterPro: IPR001788; RNA_dep_RNAPol2.
DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
KW ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
KW Transferase.
FT NP_BIND 735 742 ATP (POTENTIAL).
SQ SEQUENCE 1456 AA; 165138 MW; 10730A1EADA4FAB4 CRC64;

Query Match 4.4%: Score 102; DB 1; Length 1456;
Best Local Similarity 18.7%: Pred. No. 32;
Matches 104; Conservative 70; Mismatches 187; Indels 194; Gaps 26;

OY 52 EGLKGELQTRDVRGQATKILGVS-----ALGKLNSG-DATKIISG---CLDIYA 99

```



Db 255 SFILSNADADTRTKAAG-----VELTKVLGNWGKISQYIIAQRAAGLSTSAAGL 309  
 Qy 136 VAAYIDRLSK-----HRDEAIORHAAGAKRDPAESSAFQVWKOOSNLTDLS 185  
 Db 310 IASAVTLAISPLSLISADKFRANKRIEYSORFKKIDDDSLAAPHKEGADASLT 369  
 Qy 186 IIAANVPYKKNFNGLESRISSOGAATTSIDAKRAVDFTLLYCOLVYMRTELLVDLAI 245  
 Db 370 TI-----STVLASYSGIS-AAATTSLVGAFVSA-----LVGAVTGIIIGILE 411  
 Qy 246 LYRKGAHEVASAV-----ENANRYNKEIADTDLFLKLIPQALIGAVYHPISASETS 300  
 Db 412 ASKAMEHVASKMADVAIAEMKKKNGYFENGCDARHAAPLE-----DNF 457  
 Qy 301 KALINITYKGVDPVPRPIGNRRYKFTNSYNT 333  
 Db 458 KILSOYNKEYSV-----ERSVLITOOHMDT 482  
 RESULT 9  
 MSH5\_CAEEL STANDARD; PRT; 1369 AA.  
 ID MSH5\_CAEEL  
 AC 019272; Q9NB29;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE MSH5 protein homolog 5.  
 GN MSH-5 OR F0958.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitidae;  
 OC Rhabdilitidae; Pelodierinae; Caenorhabditis.  
 NC NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND FUNCTION.  
 RC STRAIN-Bristol N2;  
 RX MEDLINE=20469350; PubMed=11014811;  
 RA Kelly K.O., Dernburg A.F., Stanfield G.M., Villeneuve A.M.;  
 RT "Caenorhabditis elegans msh-5 is required for both normal and  
 RT radiation-induced meiotic crossing over but not for completion of  
 RT meiosis.";  
 RL Genetics 156:617-630(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Bristol N2;  
 RA Percy C.M.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP REVISIONS.  
 RA Durbin R.;  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 1-933 FROM N.A.  
 RX MEDLINE=9900534; PubMed=9787078;  
 RA Winand N.J., Panzer J.A., Kolodner R.D.;  
 RT "Cloning and characterization of the human and Caenorhabditis elegans  
 RT homologs of the Saccharomyces cerevisiae MSH5 gene.";  
 RL Genomics 53:69-80(1998).  
 CC -1- FUNCTION: Crucial component in meiotic recombination, functioning  
 CC at some point after the initiation step of recombination. Plays a  
 CC role in promoting the crossover outcome of meiotic recombination  
 CC events. Required for formation of normal meiotic crossover, and  
 CC crossover and chiasma generated by artificially made DNA breaks.  
 CC -1- SUBUNIT: HETEROOLIGOMER OF MSH4 AND MSH5 (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: Germine specific expression.  
 CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MSH5 FAMILY.  
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; AF271389; AAF76200.1;  
 DR EMBL; Z73896; CAA98059.2;  
 DR EMBL; AF070070; AAC70065.1;  
 DR WormPep; F0958.3; CE26678.  
 DR InterPro; IPR000432; Muts\_C.  
 DR InterPro; IPR002863; Muts\_N.  
 DR Pfam; PF00488; Muts\_C; 1.  
 DR Pfam; PF01624; Muts\_N; 1.  
 DR Pfam; PF01624; Muts\_C; 1.  
 DR ProDom; PD001263; Muts\_C; 1.  
 DR SMART; SM00534; Muts\_C; 1.  
 DR SMART; SM00533; Muts\_C; 1.  
 DR PROSITE; PS00486; DNA\_MISMATCH\_REPAIR\_2; 1.  
 KW Meiosis; ATP-binding; DNA-binding.  
 FT NP\_BIND 639 646 ATP (POTENTIAL).  
 SO SEQUENCE 1369 AA; 153139 MW; DDC5FDAB7DB86C4A CRC64;  
 Query Match 4.68; Score 106; DB 1; Length 1369;  
 Best Local Similarity 21.78; Pred. No. 16;  
 Matches 84; Conservative 44; Mismatches 165; Indels 94; Gaps 14;  
 Qy 35 TALKRAGSNEPALEALEGL-----KGEIQTPRVRGOATKIISV--GSLGKLN 83  
 Db 777 TVLRAGGKIKYIFRMTPLGLVDCSPALSVAKEEGIPPEVIGRACHIRYALKAGTILKEIK 836  
 Qy 84 ---SGDATKIISGCIIDYAGIATTFGPGVNGIGAVAFVSSILSLFTGSSAKNSVAVI 140  
 Db 837 AEVSNDEKOLVEDMDVVLADDEGF-----MAAVESFVKRKKSFECSRRN----V 884  
 Qy 141 DRALSKHDEAIORHAAGAKRDFEASSAFIOVMKOOSNLTDLSIIANVPYKKNFNI 200  
 Db 885 SEEIEKERSFAPSKRSRTITARSNS---VLSSRSMAVSVDQLSVLALPKKKKKVY 941  
 Qy 201 G-QLESRISSOGAATTSIDAKRAVDFTLLYCOLVYMRTELLVDLAIYRKGAHEVASAV 259  
 Db 942 GSSMSSMSPPFOEDEGTGEEDOI----SAPVSRPTL-----PSVKYASEE 987  
 Qy 260 ENANRYNKEIADTDLFLKLIP--EQALIGAVYHPISASETS----- 300  
 Db 988 EKOOSINRSRSTTAIHPIPIOMGEAGYKRRRSTSTSPGSAKSVTEVFKKTP 1047  
 Qy 301 ---KALINITYKGVDPVPRPIGNRRYKFTNSYNTSICSEAYMGNYMFGCSNVRNP 356  
 Db 1048 NVKESQVLETPPKQLSISFLEP-----KPFSEKDVLSRVSERYL----- 1087  
 Qy 357 NRVSKMSDGYTMENSDRRKLIYTK 383  
 Db 1088 -----QSDPFRT-PISDRRSQOSSRH 1107  
 RESULT 10  
 FAS3\_RHOFA STANDARD; PRT; 312 AA.  
 ID FAS3\_RHOFA  
 AC P46375;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Hypothetical 33.6 kDa protein in fasciation locus precursor (ORF3).  
 GN FAS3.  
 OS Rhodococcus fascians.  
 OC Plasmid PF1D188.  
 CC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 CC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.  
 CX NCBI\_TaxID=1628;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-D188;  
 RX MEDLINE=94222824; PubMed=8169198;  
 RA Crespi M., Vereecke D., Temmerman W., van Montagu M., Desomer J.;  
 RT "The fas operon of Rhodococcus fascians encodes new genes regulated  
 RT for efficient fasciation of host plants.";  
 RL J. Bacteriol. 176:2492-2501(1994).  
 CC -1- FUNCTION: THE FAS-OPERON ENCODES GENES INVOLVED IN CYTOKININ

FT LIPID 563 563 PALMITATE (BY SIMILARITY).  
 FT LIPID 689 689 PALMITATE (BY SIMILARITY).  
 FT VARIANT 6 6 A -> T (IN STRAIN 2001).  
 SQ SEQUENCE 1023 AA; 109867 MW; 196D5C0A9A2B5AD CRC64;  
 Query Match 4.7%; Score 108.5; DB 1; Length 1023;  
 Best Local Similarity 19.4%; Pred. No. 7.3;  
 Matches 82; Conservative 65; Mismatches 163; Indels 113; Gaps 17;  
 24 SDVNSL---LTKVETALKEASGSNEALELEGKEHQRK---DRYGOATKILGSGSA 78  
 DB 199 NNWNSFQOLNKLGSVLSNT-----KHLNKGKLNQLPNLDNIGAGLDIVSGILSA 250  
 QY 79 LGR---LNSGDA---TKIISGCLDIAGIATTEGPGMGIGAVASFVSSILSLFTGSSA 132  
 DB 251 ISASFILSNADADTGKAAAG-----VELTKYLGNNKCKISQYIIAQAAGLSLSA 305  
 QY 133 KNSVAVIDRALSK-----HDEAIQRAAGAKRDPASSAFIOVMKQOQSNLTDS 182  
 DB 306 AGLIASVWTLAISPLSELSIADKFRANKIEEYSORFKLGYDGDILAFHKEGIDA 365  
 QY 183 DSIILANYPVYKFSNFIQLESRIISOGAATSLSDAKRAVDLILYCOLVYARETLND 242  
 DB 366 SLF-----RISTVLASVSSGIS-AAATISLVGAPVSA-----LVGAVTGIIISG 407  
 QY 243 LALYKKGNAEHYASAV-----ENANRVNKLADTLDELKLIPEOCALIGAVYHPISAS 297  
 DB 408 ILEASKQAFEHYASKADYAEWKEKKGKNGYENGVDARHAFLE----- 453  
 QY 298 ETSKALANTKTYGPDVPRPIGNKRYKFTNSWNTYSISCEAYMGKNYMFRCGSNVRNP 357  
 DB 454 DNEKILSYNKEYSV-----ERSVLITQOQMDYL----- 482  
 QY 358 IRYSKMSDEFTYEMNSDR---KLYTKHDOGWGCTLEDPDQGHMRFPLRHGRKYYV 414  
 DB 483 --IGELAG---VTRNDCKILSGSYIDYEBG---KRLKKPEFOKQYFDPLKGNIDIS 534  
 QY 415 SSR 417  
 DB 535 DSK 537  
 RESULT 8  
 HLXA\_ECOLI STANDARD; PRT; 1024 AA.  
 ID HLXA\_ECOLI  
 AC P08715;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hemolysin, plasmid.  
 GN HLXA.  
 OS Escherichia coli.  
 OC Plasmid Inci2 PHLY152.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hess J., Wels W., Vogel M., Goebel W.;  
 RT "Nucleotide sequence of a plasmid-encoded hemolysin determinant and  
 its comparison with a corresponding chromosomal hemolysin sequence.";  
 RL FEMS Microbiol. Lett. 34:1-11(1986).  
 RN [2]  
 RP PALMITOYLATION OF LYS-564 AND LYS-690.  
 RX MEDLINE=95099325; PubMed=7801126;  
 RA Stanley P., Packman L.C., Koronakis V., Hughes C.;  
 RT "Fatty acylation of two internal lysine residues required for the  
 toxic activity of Escherichia coli hemolysin.";  
 RL Science 266:1992-1996(1994).  
 RN [3]  
 RP PALMITOYLATION OF LYS-564 AND LYS-690.  
 RX MEDLINE=96404790; PubMed=8808931;  
 RA Ludwig A., Garcia F., Bauer S., Jarichau T., Benz R., Hoppe J.,

RA Goebel W.;  
 RT "Analysis of the in vivo activation of hemolysin (HLXA) from  
 Escherichia coli.";  
 RL J. Bacteriol. 178:5422-5430(1996).  
 CC -1- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD  
 CC CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY  
 CC DEFINED.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING  
 CC CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC  
 CC ACTIVITY.  
 CC -1- DOMAIN: THE THREE TRANSMEMBRANE DOMAINS ARE BELIEVED TO BE  
 CC INVOLVED IN PORE FORMATION BY THE CYTOTOXIN.  
 CC -1- PTM: PALMITOYLATED BY HLXC. THE TOXIN ONLY BECOMES ACTIVE WHEN  
 CC MODIFIED.  
 CC -1- DISEASE: THE HEMOLYSIN OF E. COLI IS PRODUCED PREDOMINANTLY BY  
 CC STRAINS CAUSING EXTRAINTestinal INFECTIONS, SUCH AS THOSE OF THE  
 CC URINARY TRACT.  
 CC -1- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-slb.ch/announce/>  
 CC or send an email to [license@slb.ch](mailto:license@slb.ch)).  
 CC  
 DR EMBL: M14107; AAA98233.1;  
 DR InterPro: IPR001343; Hemolysin\_Ca\_bind.  
 DR InterPro: IPR003355; RTXtoxin\_N.  
 DR InterPro: IPR003995; RtxA.  
 DR Pfam: PF00353; hemolysinCbind; 6.  
 DR Pfam: PF02382; RTX; 1.  
 DR PRINTS: PR00313; CABDNCRPT.  
 DR PRINTS: PR01488; RTXTOXINA.  
 DR PROSITE: PS00330; HEMOLYSIN\_CALCIUM; 4.  
 DR Hemolysin; Toxin; Cytolysin; Cytotoxin; Repeat; Calcium;  
 KW Transmembrane; Lipoprotein; Palmitate; Plasmid.  
 FT TRANSMEM 238 260  
 FT TRANSMEM 268 327  
 FT TRANSMEM 365 411  
 FT DOMAIN 724 870  
 FT REPEAT 724 729  
 FT REPEAT 723 738  
 FT REPEAT 742 747  
 FT REPEAT 751 756  
 FT REPEAT 760 765  
 FT REPEAT 769 774  
 FT REPEAT 778 783  
 FT REPEAT 787 792  
 FT REPEAT 796 801  
 FT REPEAT 807 812  
 FT REPEAT 817 822  
 FT REPEAT 826 831  
 FT REPEAT 835 840  
 FT REPEAT 844 849  
 FT REPEAT 856 861  
 FT REPEAT 865 870  
 FT LIPID 564 564  
 FT LIPID 564 564  
 FT LIPID 564 564  
 FT LIPID 564 564  
 SQ SEQUENCE 1024 AA; 110201 MW; 83944917F76C945B CRC64;  
 Query Match 4.7%; Score 108.5; DB 1; Length 1024;  
 Best Local Similarity 20.7%; Pred. No. 7.3;  
 Matches 69; Conservative 52; Mismatches 139; Indels 73; Gaps 12;  
 24 SDVNSLTKVETALKEASGSNEALELEGKEHQRK---DRYGOATKILGSGALGK 81  
 DB 200 NNWNSFQOLNKLGSVLSNT-----KHLNKGKLNQLPNLDNIGAGLDIVSGILSA 254  
 QY 82 ---LNSGDA---TKIISGCLDIAGIATTEGPGMGIGAVASFVSSILSLFTGSSAKNS 135

MultiGene family: Hypothetical protein.  
 KM DOMAIN 1 181 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 182 202 POTENTIAL.  
 FT DOMAIN 203 220 LUMENAL (POTENTIAL).  
 FT TRANSMEM 221 241 POTENTIAL.  
 FT DOMAIN 242 261 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 262 289 POTENTIAL.  
 FT DOMAIN 290 309 LUMENAL (POTENTIAL).  
 FT TRANSMEM 310 326 POTENTIAL.  
 FT DOMAIN 327 344 POTENTIAL.  
 FT TRANSMEM 345 363 POTENTIAL.  
 FT DOMAIN 364 384 LUMENAL (POTENTIAL).  
 FT TRANSMEM 385 395 POTENTIAL.  
 FT DOMAIN 396 415 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 416 438 POTENTIAL.  
 FT DOMAIN 439 458 LUMENAL (POTENTIAL).  
 FT TRANSMEM 459 477 POTENTIAL.  
 FT DOMAIN 478 497 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 498 516 POTENTIAL.  
 FT DOMAIN 517 536 LUMENAL (POTENTIAL).  
 FT TRANSMEM 537 555 POTENTIAL.  
 FT DOMAIN 556 574 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 575 593 POTENTIAL.  
 FT DOMAIN 594 612 LUMENAL (POTENTIAL).  
 FT TRANSMEM 613 631 POTENTIAL.  
 FT DOMAIN 632 650 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 651 669 POTENTIAL.  
 FT DOMAIN 670 688 LUMENAL (POTENTIAL).  
 FT TRANSMEM 689 707 POTENTIAL.  
 FT DOMAIN 708 726 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 727 745 POTENTIAL.  
 FT DOMAIN 746 764 LUMENAL (POTENTIAL).  
 FT TRANSMEM 765 783 POTENTIAL.  
 FT DOMAIN 784 802 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 803 821 POTENTIAL.  
 FT DOMAIN 822 840 LUMENAL (POTENTIAL).  
 FT TRANSMEM 841 859 POTENTIAL.  
 FT DOMAIN 860 878 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 879 897 POTENTIAL.  
 FT DOMAIN 898 916 LUMENAL (POTENTIAL).  
 FT TRANSMEM 917 935 POTENTIAL.  
 FT DOMAIN 936 954 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 955 973 POTENTIAL.  
 FT DOMAIN 974 992 LUMENAL (POTENTIAL).  
 FT TRANSMEM 993 1011 POTENTIAL.  
 FT DOMAIN 1012 1030 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 1031 1049 POTENTIAL.  
 FT DOMAIN 1050 1068 LUMENAL (POTENTIAL).  
 FT TRANSMEM 1069 1087 POTENTIAL.  
 FT DOMAIN 1088 1106 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 1107 1125 POTENTIAL.  
 FT DOMAIN 1126 1144 LUMENAL (POTENTIAL).  
 FT TRANSMEM 1145 1163 POTENTIAL.  
 FT DOMAIN 1164 1182 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 1183 1201 POTENTIAL.  
 FT DOMAIN 1202 1220 LUMENAL (POTENTIAL).  
 FT TRANSMEM 1221 1239 POTENTIAL.  
 FT DOMAIN 1240 1258 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 1259 1277 POTENTIAL.  
 FT DOMAIN 1278 1296 LUMENAL (POTENTIAL).  
 FT TRANSMEM 1297 1315 POTENTIAL.  
 FT DOMAIN 1316 1334 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 1335 1353 POTENTIAL.  
 FT DOMAIN 1354 1372 LUMENAL (POTENTIAL).  
 FT TRANSMEM 1373 1391 POTENTIAL.  
 FT DOMAIN 1392 1410 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 1411 1429 POTENTIAL.  
 FT DOMAIN 1430 1448 LUMENAL (POTENTIAL).  
 FT TRANSMEM 1449 1467 POTENTIAL.  
 FT DOMAIN 1468 1486 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 1487 1505 POTENTIAL.  
 FT DOMAIN 1506 1524 LUMENAL (POTENTIAL).  
 FT TRANSMEM 1525 1543 POTENTIAL.  
 FT DOMAIN 1544 1562 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 1563 1581 POTENTIAL.  
 FT DOMAIN 1582 1600 LUMENAL (POTENTIAL).  
 FT TRANSMEM 1601 1619 POTENTIAL.  
 FT DOMAIN 1620 1638 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 1639 1657 POTENTIAL.  
 FT DOMAIN 1658 1676 LUMENAL (POTENTIAL).  
 FT TRANSMEM 1677 1695 POTENTIAL.  
 FT DOMAIN 1696 1714 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 1715 1733 POTENTIAL.  
 FT DOMAIN 1734 1752 LUMENAL (POTENTIAL).  
 FT TRANSMEM 1753 1771 POTENTIAL.  
 FT DOMAIN 1772 1790 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 1791 1809 POTENTIAL.  
 FT DOMAIN 1810 1828 LUMENAL (POTENTIAL).  
 FT TRANSMEM 1829 1847 POTENTIAL.  
 FT DOMAIN 1848 1866 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 1867 1885 POTENTIAL.  
 FT DOMAIN 1886 1904 LUMENAL (POTENTIAL).  
 FT TRANSMEM 1905 1923 POTENTIAL.  
 FT DOMAIN 1924 1942 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 1943 1961 POTENTIAL.  
 FT DOMAIN 1962 1980 LUMENAL (POTENTIAL).  
 FT TRANSMEM 1981 1999 POTENTIAL.  
 FT DOMAIN 2000 2018 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 2019 2037 POTENTIAL.  
 FT DOMAIN 2038 2056 LUMENAL (POTENTIAL).  
 FT TRANSMEM 2057 2075 POTENTIAL.  
 FT DOMAIN 2076 2094 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 2095 2113 POTENTIAL.  
 FT DOMAIN 2114 2132 LUMENAL (POTENTIAL).  
 FT TRANSMEM 2133 2151 POTENTIAL.  
 FT DOMAIN 2152 2170 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 2171 2189 POTENTIAL.  
 FT DOMAIN 2190 2208 LUMENAL (POTENTIAL).  
 FT TRANSMEM 2209 2227 POTENTIAL.  
 FT DOMAIN 2228 2246 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 2247 2265 POTENTIAL.  
 FT DOMAIN 2266 2284 LUMENAL (POTENTIAL).  
 FT TRANSMEM 2285 2303 POTENTIAL.  
 FT DOMAIN 2304 2322 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 2323 2341 POTENTIAL.  
 FT DOMAIN 2342 2360 LUMENAL (POTENTIAL).  
 FT TRANSMEM 2361 2379 POTENTIAL.  
 FT DOMAIN 2380 2398 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 2399 2417 POTENTIAL.  
 FT DOMAIN 2418 2436 LUMENAL (POTENTIAL).  
 FT TRANSMEM 2437 2455 POTENTIAL.  
 FT DOMAIN 2456 2474 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 2475 2493 POTENTIAL.  
 FT DOMAIN 2494 2512 LUMENAL (POTENTIAL).  
 FT TRANSMEM 2513 2531 POTENTIAL.  
 FT DOMAIN 2532 2550 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 2551 2569 POTENTIAL.  
 FT DOMAIN 2570 2588 LUMENAL (POTENTIAL).  
 FT TRANSMEM 2589 2607 POTENTIAL.  
 FT DOMAIN 2608 2626 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 2627 2645 POTENTIAL.  
 FT DOMAIN 2646 2664 LUMENAL (POTENTIAL).  
 FT TRANSMEM 2665 2683 POTENTIAL.  
 FT DOMAIN 2684 2702 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 2703 2721 POTENTIAL.  
 FT DOMAIN 2722 2740 LUMENAL (POTENTIAL).  
 FT TRANSMEM 2741 2759 POTENTIAL.  
 FT DOMAIN 2760 2778 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 2779 2797 POTENTIAL.  
 FT DOMAIN 2798 2816 LUMENAL (POTENTIAL).  
 FT TRANSMEM 2817 2835 POTENTIAL.  
 FT DOMAIN 2836 2854 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 2855 2873 POTENTIAL.  
 FT DOMAIN 2874 2892 LUMENAL (POTENTIAL).  
 FT TRANSMEM 2893 2911 POTENTIAL.  
 FT DOMAIN 2912 2930 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 2931 2949 POTENTIAL.  
 FT DOMAIN 2950 2968 LUMENAL (POTENTIAL).  
 FT TRANSMEM 2969 2987 POTENTIAL.  
 FT DOMAIN 2988 3006 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 3007 3025 POTENTIAL.  
 FT DOMAIN 3026 3044 LUMENAL (POTENTIAL).  
 FT TRANSMEM 3045 3063 POTENTIAL.  
 FT DOMAIN 3064 3082 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 3083 3101 POTENTIAL.  
 FT DOMAIN 3102 3120 LUMENAL (POTENTIAL).  
 FT TRANSMEM 3121 3139 POTENTIAL.  
 FT DOMAIN 3140 3158 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 3159 3177 POTENTIAL.  
 FT DOMAIN 3178 3196 LUMENAL (POTENTIAL).  
 FT TRANSMEM 3197 3215 POTENTIAL.  
 FT DOMAIN 3216 3234 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 3235 3253 POTENTIAL.  
 FT DOMAIN 3254 3272 LUMENAL (POTENTIAL).  
 FT TRANSMEM 3273 3291 POTENTIAL.  
 FT DOMAIN 3292 3310 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 3311 3329 POTENTIAL.  
 FT DOMAIN 3330 3348 LUMENAL (POTENTIAL).  
 FT TRANSMEM 3349 3367 POTENTIAL.  
 FT DOMAIN 3368 3386 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 3387 3405 POTENTIAL.  
 FT DOMAIN 3406 3424 LUMENAL (POTENTIAL).  
 FT TRANSMEM 3425 3443 POTENTIAL.  
 FT DOMAIN 3444 3462 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 3463 3481 POTENTIAL.  
 FT DOMAIN 3482 3500 LUMENAL (POTENTIAL).  
 FT TRANSMEM 3501 3519 POTENTIAL.  
 FT DOMAIN 3520 3538 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 3539 3557 POTENTIAL.  
 FT DOMAIN 3558 3576 LUMENAL (POTENTIAL).  
 FT TRANSMEM 3577 3595 POTENTIAL.  
 FT DOMAIN 3596 3614 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 3615 3633 POTENTIAL.  
 FT DOMAIN 3634 3652 LUMENAL (POTENTIAL).  
 FT TRANSMEM 3653 3671 POTENTIAL.  
 FT DOMAIN 3672 3690 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 3691 3709 POTENTIAL.  
 FT DOMAIN 3710 3728 LUMENAL (POTENTIAL).  
 FT TRANSMEM 3729 3747 POTENTIAL.  
 FT DOMAIN 3748 3766 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 3767 3785 POTENTIAL.  
 FT DOMAIN 3786 3804 LUMENAL (POTENTIAL).  
 FT TRANSMEM 3805 3823 POTENTIAL.  
 FT DOMAIN 3824 3842 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 3843 3861 POTENTIAL.  
 FT DOMAIN 3862 3880 LUMENAL (POTENTIAL).  
 FT TRANSMEM 3881 3899 POTENTIAL.  
 FT DOMAIN 3900 3918 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 3919 3937 POTENTIAL.  
 FT DOMAIN 3938 3956 LUMENAL (POTENTIAL).  
 FT TRANSMEM 3957 3975 POTENTIAL.  
 FT DOMAIN 3976 3994 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 3995 4013 POTENTIAL.  
 FT DOMAIN 4014 4032 LUMENAL (POTENTIAL).  
 FT TRANSMEM 4033 4051 POTENTIAL.  
 FT DOMAIN 4052 4070 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 4071 4089 POTENTIAL.  
 FT DOMAIN 4090 4108 LUMENAL (POTENTIAL).  
 FT TRANSMEM 4109 4127 POTENTIAL.  
 FT DOMAIN 4128 4146 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 4147 4165 POTENTIAL.  
 FT DOMAIN 4166 4184 LUMENAL (POTENTIAL).  
 FT TRANSMEM 4185 4203 POTENTIAL.  
 FT DOMAIN 4204 4222 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 4223 4241 POTENTIAL.  
 FT DOMAIN 4242 4260 LUMENAL (POTENTIAL).  
 FT TRANSMEM 4261 4279 POTENTIAL.  
 FT DOMAIN 4280 4298 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 4299 4317 POTENTIAL.  
 FT DOMAIN 4318 4336 LUMENAL (POTENTIAL).  
 FT TRANSMEM 4337 4355 POTENTIAL.  
 FT DOMAIN 4356 4374 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 4375 4393 POTENTIAL.  
 FT DOMAIN 4394 4412 LUMENAL (POTENTIAL).  
 FT TRANSMEM 4413 4431 POTENTIAL.  
 FT DOMAIN 4432 4450 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 4451 4469 POTENTIAL.  
 FT DOMAIN 4470 4488 LUMENAL (POTENTIAL).  
 FT TRANSMEM 4489 4507 POTENTIAL.  
 FT DOMAIN 4508 4526 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 4527 4545 POTENTIAL.  
 FT DOMAIN 4546 4564 LUMENAL (POTENTIAL).  
 FT TRANSMEM 4565 4583 POTENTIAL.  
 FT DOMAIN 4584 4602 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 4603 4621 POTENTIAL.  
 FT DOMAIN 4622 4640 LUMENAL (POTENTIAL).  
 FT TRANSMEM 4641 4659 POTENTIAL.  
 FT DOMAIN 4660 4678 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 4679 4697 POTENTIAL.  
 FT DOMAIN 4698 4716 LUMENAL (POTENTIAL).  
 FT TRANSMEM 4717 4735 POTENTIAL.  
 FT DOMAIN 4736 4754 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 4755 4773 POTENTIAL.  
 FT DOMAIN 4774 4792 LUMENAL (POTENTIAL).  
 FT TRANSMEM 4793 4811 POTENTIAL.  
 FT DOMAIN 4812 4830 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 4831 4849 POTENTIAL.  
 FT DOMAIN 4850 4868 LUMENAL (POTENTIAL).  
 FT TRANSMEM 4869 4887 POTENTIAL.  
 FT DOMAIN 4888 4906 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 4907 4925 POTENTIAL.  
 FT DOMAIN 4926 4944 LUMENAL (POTENTIAL).  
 FT TRANSMEM 4945 4963 POTENTIAL.  
 FT DOMAIN 4964 4982 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 4983 5001 POTENTIAL.  
 FT DOMAIN 5002 5020 LUMENAL (POTENTIAL).  
 FT TRANSMEM 5021 5039 POTENTIAL.  
 FT DOMAIN 5040 5058 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 5059 5077 POTENTIAL.  
 FT DOMAIN 5078 5096 LUMENAL (POTENTIAL).  
 FT TRANSMEM 5097 5115 POTENTIAL.  
 FT DOMAIN 5116 5134 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 5135 5153 POTENTIAL.  
 FT DOMAIN 5154 5172 LUMENAL (POTENTIAL).  
 FT TRANSMEM 5173 5191 POTENTIAL.  
 FT DOMAIN 5192 5210 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 5211 5229 POTENTIAL.  
 FT DOMAIN 5230 5248 LUMENAL (POTENTIAL).  
 FT TRANSMEM 5249 5267 POTENTIAL.  
 FT DOMAIN 5268 5286 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 5287 5305 POTENTIAL.  
 FT DOMAIN 5306 5324 LUMENAL (POTENTIAL).  
 FT TRANSMEM 5325 5343 POTENTIAL.  
 FT DOMAIN 5344 5362 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 5363 5381 POTENTIAL.  
 FT DOMAIN 5382 5400 LUMENAL (POTENTIAL).  
 FT TRANSMEM 5401 5419 POTENTIAL.  
 FT DOMAIN 5420 5438 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 5439 5457 POTENTIAL.  
 FT DOMAIN 5458 5476 LUMENAL (POTENTIAL).  
 FT TRANSMEM 5477 5495 POTENTIAL.  
 FT DOMAIN 5496 5514 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 5515 5533 POTENTIAL.  
 FT DOMAIN 5534 5552 LUMENAL (POTENTIAL).  
 FT TRANSMEM 5553 5571 POTENTIAL.  
 FT DOMAIN 5572 5590 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 5591 5609 POTENTIAL.  
 FT DOMAIN 5610 5628 LUMENAL (POTENTIAL).  
 FT TRANSMEM 5629 5647 POTENTIAL.  
 FT DOMAIN 5648 5666 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 5667 5685 POTENTIAL.  
 FT DOMAIN 5686 5704 LUMENAL (POTENTIAL).  
 FT TRANSMEM 5705 5723 POTENTIAL.  
 FT DOMAIN 5724 5742 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 5743 5761 POTENTIAL.  
 FT DOMAIN 5762 5780 LUMENAL (POTENTIAL).  
 FT TRANSMEM 5781 5800 POTENTIAL.  
 FT DOMAIN 5801 5818 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 5819 5837 POTENTIAL.  
 FT DOMAIN 5838 5856 LUMENAL (POTENTIAL).  
 FT TRANSMEM 5857 5875 POTENTIAL.  
 FT DOMAIN 5876 5894 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 5895 5913 POTENTIAL.  
 FT DOMAIN 5914 5932 LUMENAL (POTENTIAL).  
 FT TRANSMEM 5933 5951 POTENTIAL.  
 FT DOMAIN 5952 5970 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 5971 5989 POTENTIAL.  
 FT DOMAIN 5990 6008 LUMENAL (POTENTIAL).  
 FT TRANSMEM 6009 6027 POTENTIAL.  
 FT DOMAIN 6028 6046 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 6047 6065 POTENTIAL.  
 FT DOMAIN 6066 6084 LUMENAL (POTENTIAL).  
 FT TRANSMEM 6085 6103 POTENTIAL.  
 FT DOMAIN 6104 6122 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 6123 6141 POTENTIAL.  
 FT DOMAIN 6142 6160 LUMENAL (POTENTIAL).  
 FT TRANSMEM 6161 6179 POTENTIAL.  
 FT DOMAIN 6180 6198 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 6199 6217 POTENTIAL.  
 FT DOMAIN 6218 6236 LUMENAL (POTENTIAL).  
 FT TRANSMEM 6237 6255 POTENTIAL.  
 FT DOMAIN 6256 6274 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 6275 6293 POTENTIAL.  
 FT DOMAIN 6294 6312 LUMENAL (POTENTIAL).  
 FT TRANSMEM 6313 6331 POTENTIAL.  
 FT DOMAIN 6332 6350 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 6351 6369 POTENTIAL.  
 FT DOMAIN 6370 6388 LUMENAL (POTENTIAL).  
 FT TRANSMEM 6389 6407 POTENTIAL.  
 FT DOMAIN 6408 6426 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 6427 6445 POTENTIAL.  
 FT DOMAIN 6446 6464 LUMENAL (POTENTIAL).  
 FT TRANSMEM 6465 6483 POTENTIAL.  
 FT DOMAIN 6484 6502 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 6503 6521 POTENTIAL.  
 FT DOMAIN 6522 6540 LUMENAL (POTENTIAL).  
 FT TRANSMEM 6541 6559 POTENTIAL.  
 FT DOMAIN 6560 6578 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 6579 6597 POTENTIAL.  
 FT DOMAIN 6598 6616 LUMENAL (POTENTIAL).  
 FT TRANSMEM 6617 6635 POTENTIAL.  
 FT DOMAIN 6636 6654 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 6655 6673 POTENTIAL.  
 FT DOMAIN 6674 6692 LUMENAL (POTENTIAL).  
 FT TRANSMEM 6693 6711 POTENTIAL.  
 FT DOMAIN 6712 6730 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 6731 6749 POTENTIAL.  
 FT DOMAIN 6750 6768 LUMENAL (POTENTIAL).  
 FT TRANSMEM 6769 6787 POTENTIAL.  
 FT DOMAIN 6788 6806 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 6807 6825 POTENTIAL.  
 FT DOMAIN 6826 6844 LUMENAL (POTENTIAL).  
 FT TRANSMEM 6845 6863 POTENTIAL.  
 FT DOMAIN 6864 6882 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 6883 6901 POTENTIAL.  
 FT DOMAIN 6902 6920 LUMENAL (POTENTIAL).  
 FT TRANSMEM 6921 6939 POTENTIAL.  
 FT DOMAIN 6940 6958 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 6959 6977 POTENTIAL.  
 FT DOMAIN 6978 6996 LUMENAL (POTENTIAL).  
 FT TRANSMEM 6997 7015 POTENTIAL.  
 FT DOMAIN 7016 7034 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 7035 7053 POTENTIAL.  
 FT DOMAIN 7054 7072 LUMENAL (POTENTIAL).  
 FT TRANSMEM 7073 7091 POTENTIAL.  
 FT DOMAIN 7092 7110 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 7111 7129 POTENTIAL.  
 FT DOMAIN 7130 7148 LUMENAL (POTENTIAL).  
 FT TRANSMEM 7149 7167 POTENTIAL.  
 FT DOMAIN 7168 7186 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 7187 7205 POTENTIAL.  
 FT DOMAIN 7206 7224 LUMENAL (POTENTIAL).  
 FT TRANSMEM 7225 7243 POTENTIAL.  
 FT DOMAIN 7244 7262 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 7263 7281 POTENTIAL.  
 FT DOMAIN 7282 7300 LUMENAL (POTENTIAL).  
 FT TRANSMEM 7301 7319 POTENTIAL.  
 FT DOMAIN 7320 7338 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 7339 7357 POTENTIAL.  
 FT DOMAIN 7358 7376 LUMENAL (POTENTIAL).  
 FT TRANSMEM 7377 7395 POTENTIAL.  
 FT DOMAIN 7396 7414 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 7415 7433 POTENTIAL.  
 FT DOMAIN 7434 7452 LUMENAL (POTENTIAL).  
 FT TRANSMEM 7453 7471 POTENTIAL.  
 FT DOMAIN 7472 7490 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 7491 7509 POTENTIAL.  
 FT DOMAIN 7510 7528 LUMENAL (POTENTIAL).  
 FT TRANSMEM 7529 7547 POTENTIAL.  
 FT DOMAIN 7548 7566 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 7567 7585 POTENTIAL.  
 FT DOMAIN 7586 7604 LUMENAL (POTENTIAL).  
 FT TRANSMEM 7605 7623 POTENTIAL.  
 FT DOMAIN 7624 7642 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 7643 7661 POTENTIAL.  
 FT DOMAIN 7662 7680 LUMENAL (POTENTIAL).  
 FT TRANSMEM 7681 7699 POTENTIAL.  
 FT DOMAIN 7700 7718 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 7719 7737 POTENTIAL.  
 FT DOMAIN 7738 7756 LUMENAL (POTENTIAL).  
 FT TRANSMEM 7757 7775 POTENTIAL.  
 FT DOMAIN 7776 7794 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 7795 7813 POTENTIAL.  
 FT DOMAIN 7814 7832 LUMENAL (POTENTIAL).  
 FT TRANSMEM 7833 7851 POTENTIAL.  
 FT DOMAIN 7852 7870 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 7871 7889 POTENTIAL.  
 FT DOMAIN 7890 7908 LUMENAL (POTENTIAL).  
 FT TRANSMEM 7909 7927 POTENTIAL.  
 FT DOMAIN 7928 7946 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 7947 7965 POTENTIAL.  
 FT DOMAIN 7966 7984 LUMENAL (POTENTIAL).  
 FT TRANSMEM 7985 8003 POTENTIAL.  
 FT DOMAIN 8004 8022 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 8023 8041 POTENTIAL.  
 FT DOMAIN 8042 8060 LUMENAL (POTENTIAL).  
 FT TRANSMEM 8061 8079 POTENTIAL.  
 FT DOMAIN 8080 8108 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 8109 8127 POTENTIAL.  
 FT DOMAIN 8128 8146 LUMENAL (POTENTIAL).  
 FT TRANSMEM 8147 8165 POTENTIAL.  
 FT DOMAIN 8166 8184 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 8185 8203 POTENTIAL.  
 FT DOMAIN 8204 8222 LUMENAL (POTENTIAL).  
 FT TRANSMEM 8223 8241 POTENTIAL.  
 FT DOMAIN 8242 8260 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 8261 8279 POTENTIAL.  
 FT DOMAIN 8280 8308 LUMENAL (POTENTIAL).  
 FT TRANSMEM 8309 8327 POTENTIAL.  
 FT DOMAIN 8328 8346 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 8347 8365 POTENTIAL.  
 FT DOMAIN 8366 8384 LUMENAL (POTENTIAL).  
 FT TRANSMEM 8385 8403 POTENTIAL.  
 FT DOMAIN 8404 8422 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 8423 8441 POTENTIAL.  
 FT DOMAIN 8442 8460 LUMENAL (POTENTIAL).  
 FT TRANSMEM 8461 8479 POTENTIAL.  
 FT DOMAIN 8480 8498 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 8499 8517 POTENTIAL.  
 FT DOMAIN 8518 8536 LUMENAL (POTENTIAL).  
 FT TRANSMEM 8537 8555 POTENTIAL.  
 FT DOMAIN 8556 8574 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 8575 8593 POTENTIAL.  
 FT DOMAIN 8594 8612 LUMENAL (POTENTIAL).  
 FT TRANSMEM 8613 8631 POTENTIAL.  
 FT DOMAIN 8632 8650 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 8651 8669 POTENTIAL.  
 FT DOMAIN 8670 8688 LUMENAL (POTENTIAL).  
 FT TRANSMEM 8689 8707 POTENTIAL.  
 FT DOMAIN 8708 8726 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 8727 8745 POTENTIAL.  
 FT DOMAIN 8746 8764 LUMENAL (POTENTIAL).  
 FT TRANSMEM 8765 8783 POTENTIAL.  
 FT DOMAIN 8784 8802 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 8803 8821 POTENTIAL.  
 FT DOMAIN 8822 8840 LUMENAL (POTENTIAL).  
 FT TRANSMEM 8841 8859 POTENTIAL.  
 FT DOMAIN 8860 8878 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 8879 8897 POTENTIAL.  
 FT DOMAIN 8898 8916 LUMENAL (POTENTIAL).  
 FT TRANSMEM 8917 8935 POTENTIAL.  
 FT DOMAIN 8936 8954 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 8955 8973 POTENTIAL.  
 FT DOMAIN 8974 8992 LUMENAL (POTENTIAL).  
 FT TRANSMEM 8993 9011 POTENTIAL.  
 FT DOMAIN 9012 9030 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 9031 9049 POTENTIAL.  
 FT DOMAIN 9050 9068 LUMENAL (POTENTIAL).  
 FT TRANSMEM 9069 9087 POTENTIAL.  
 FT DOMAIN 9088 9106 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 9107 9125 POTENTIAL.  
 FT DOMAIN 9126 9144 LUMENAL (POTENTIAL).  
 FT TRANSMEM 9145 9163 POTENTIAL.  
 FT DOMAIN 9164 9182 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 9183 9201 POTENTIAL.  
 FT DOMAIN 9202 9220 LUMENAL (POTENTIAL).  
 FT TRANSMEM 9221 9239 POTENTIAL.  
 FT DOMAIN 9240 9258 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 9259 9277 POTENTIAL.  
 FT DOMAIN 9278 9296 LUMENAL (POTENTIAL).  
 FT TRANSMEM 9297 9315 POTENTIAL.  
 FT DOMAIN 9316 9334 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 9335 9353 POTENTIAL.  
 FT DOMAIN 9354 9372 LUMENAL (POTENTIAL).  
 FT TRANSMEM 9373 9391 POTENTIAL.  
 FT DOMAIN 9392 9410 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 9411 9429 POTENTIAL.  
 FT DOMAIN 9430 9448 LUMENAL (POTENTIAL).  
 FT TRANSMEM 9449 9467 POTENTIAL.  
 FT DOMAIN 9468 9486 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 9487 9505 POTENTIAL.  
 FT DOMAIN 9506 9524 LUMENAL (POTENTIAL).  
 FT TRANSMEM 9525 9543 POTENTIAL.  
 FT DOMAIN 9544 9562 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 9563 9581 POTENTIAL.  
 FT DOMAIN 9582 9600 LUMENAL (POTENTIAL).  
 FT TRANSMEM 9601 9619 POTENTIAL.  
 FT DOMAIN 9620 9638 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 9639 9657 POTENTIAL.  
 FT DOMAIN 9658 9676 LUMENAL (POTENTIAL).  
 FT TRANSMEM 9677 9695 POTENTIAL.  
 FT DOMAIN 9696 9714 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 9715 9733 POTENTIAL.  
 FT DOMAIN 9734 9752 LUMENAL (POTENTIAL).  
 FT TRANSMEM 9753 9771 POTENTIAL.  
 FT DOMAIN 9772 9790 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 9791 9809 POTENTIAL.  
 FT DOMAIN 9810 9828 LUMENAL (POTENTIAL).  
 FT TRANSMEM 9829 9847 POTENTIAL.  
 FT DOMAIN 9848 9866 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 9867 9885 POTENTIAL.  
 FT DOMAIN 9886 9904 LUMENAL (POTENTIAL).  
 FT TRANSMEM 9905 9923 POTENTIAL.  
 FT DOMAIN 9924 9942 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 9943 9961 POTENTIAL.  
 FT DOMAIN 9962 9980 LUMENAL (POTENTIAL).  
 FT TRANSMEM 9981 9999 POTENTIAL.

Query Match 4.7%; Score 110; DB 1; Length 10







Db 61 KPRVQATKILGVSALGKLSGDAITIIISGCLDIIVAGIATTEGCPVGIGIAVASFV 120  
 QY 121 SSILSFTGSSAKNSVAVIDRALSKRDEAIORHAGAKRDEFAESSAFIOVMQOENLT 180  
 Db 121 SSILSFTGSSAKNSVAVIDRALSKRDEAIORHAGAKRDEFAESSAFIOVMQOENLT 180  
 QY 181 DSDLSIIAANVPYKFNFGOLESRISOGAATTSLSDAKRAVDIFILLYCOLVYMRRL 240  
 Db 181 DSDLSIIAANVPYKFNFGOLESRISOGAATTSLSDAKRAVDIFILLYCOLVYMRRL 240  
 QY 241 VDIALIYRKGNAEHVAHAVENANVNKELADTIDFLHKLIPDALIGAYHPIASETS 300  
 Db 241 VDIALIYRKGNAEHVAHAVENANVNKELADTIDFLHKLIPDALIGAYHPIASETS 300  
 QY 301 KALINTKTYGVDPVPRPIGNRRYKFTNSYMTYSISCEAYMGNYMFGCSNVNPNIRV 360  
 Db 301 KALINTKTYGVDPVPRPIGNRRYKFTNSYMTYSISCEAYMGNYMFGCSNVNPNIRV 360  
 QY 361 SKMSDGYTWMNSDRRLYITKHDGCGWGTLDDEPDGCHMRFIPLRHGKYMVSSKRW 420  
 Db 361 SKMSDGYTWMNSDRRLYITKHDGCGWGTLDDEPDGCHMRFIPLRHGKYMVSSKRW 420  
 QY 421 NMFMYESSASGYIRSMENNPPOGHWSIT 450  
 Db 421 NMFMYESSASGYIRSMENNPPOGHWSIT 450

## RESULT 2

CTXA\_CARAL STANDARD: PRT: 463 AA.

AC Q9GN8; 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Toxin-A precursor (CatX-A).  
 GN CAT-1.  
 OS Carybdea alata (Hawaiian box jellyfish).  
 OC Eukaryota; Metazoa; Cnidaria; Cubozoa; Cubomedusae; Carybdeidae;  
 OC Carybdea.  
 OX NCBI\_TaxID-112899;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 158-168 AND 431-448.  
 RC TISSUE=Cnidoblast;  
 RX MEDLINE=20422302; PubMed=10964708;  
 RA Nagai H., Takawa K., Nakao M., Sakamoto B., Crow G.L., Nakajima T.;  
 RT "Isolation and characterization of a novel protein toxin from the  
 Hawaiian box jellyfish (sea wasp) Carybdea alata.";  
 RL Biochem. Biophys. Res. Commun. 275:589-594(2000).  
 CC -1- FUNCTION: Has potent hemolytic activity. lethally toxic to  
 crayfish when administered via intraperitoneal injection (LD50 =  
 5-25 microg/kg). Causes cutaneous inflammation in humans.  
 CC -1- PTM: There are disulfide bonds.  
 CC -1- SIMILARITY: BELONGS TO THE JELLYFISH TOXIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB036714; BAB12727.1;  
 KW Toxin; Hemolysis; Signal.  
 FT SIGNAL 1 18  
 CHAIN 19 463  
 SEQUENCE 463 AA: 51604 MW: 461253DFCEB3D3C0 CRC64:

Query Match 42.7%; Score 992.5; DB 1; Length 463;  
 Best Local Similarity 43.5%; Pred. No. 2.2e-61;  
 Matches 197; Conservative 95; Mismatches 148; Indels 13; Gaps 7;

QY 9 LFTVLAITSARKG-----KRSVDNSLITKVELTALKRAS--GSNEAL-EALEGLKEIQT 60  
 Db 11 LFTVLAITSARKG-----KRSVDNSLITKVELTALKRAS--GSNEAL-EALEGLKEIQT 60  
 QY 61 KP-DRGQATKILGVSALGKLSGDAITIIISGCLDIIVAGIATTEGCPVGIGIAVASFV 119  
 Db 71 EPATTTAKVSTIVGSGVSLSKFSGDPPEVAGSCIDIIVATFTGPGYGAIAVASTL 130  
 QY 120 VSSILSFTGSSAKNSVAVIDRALSKRDEAIORHAGAKRDEFAESSAFIOVMQOENLT 179  
 Db 131 SSILSFTGSSAKNSVAVIDRALSKRDEAIORHAGAKRDEFAESSAFIOVMQOENLT 190  
 QY 180 TDSLSIIAANVPYKFNFGOLESRISOGAATTSLSDAKRAVDIFILLYCOLVYMRRL 239  
 Db 191 TEVHDSVDAVAVDAFTNMLGVLSEIRINRGSTVDNNEAMKRTNIFPLQLSVARETL 250  
 QY 240 LVDLAILYRK--GNAEHVAHAVENANVNKELADTIDFLHKLIPDALIGAYHPIAS 297  
 Db 251 LTVQVILLYRAGAYDELALSLISLSDQNKKEATRETFTLHQMETKYSICGSYYPIIDS 310  
 QY 298 ESKALINTKTYGVDPVPRPIGNRRYKFTNSYMTYSISCEAYMGNYMFGCSNVNPNIRV 356  
 Db 311 KALIGILKLTKEFGVDPARYTFDGLYRMQNANMRSICKESYAGNHFRGCKXDSYH 370  
 QY 357 NTRVSKMSDGYTWMNSDRRLYITKHDGCGWGTLDDEPDGCHMRFIPLRHGKYMVSS 416  
 Db 371 GIRIKLENGYHTI-ILSKAMTVTKHAGCGWGTLDDEPDGCHMRFIPLRHGKYMVSS 429  
 QY 417 KRPNMFMYMESSASGYIRSMENNPPOGHWSI 449  
 Db 430 KRPNMFMYMESSASGYIRSMENNPPOGHWSI 462

## RESULT 3

CTXA\_CHIOU STANDARD: PRT: 462 AA.

AC P58762; 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Toxin-A precursor (CqTx-A).  
 GN CQT-A.  
 OS Chirospalmus quadrangatus (Box jellyfish).  
 OC Eukaryota; Metazoa; Cnidaria; Cubozoa; Cubomedusae; Chirodroplidae;  
 OC Chirospalmus.  
 OX NCBI\_TaxID-130731;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Cnidoblast;  
 RX MEDLINE=21854932; PubMed=11866126;  
 RA Nagai H., Takawa K., Nakao M., Oshiro N., Iwanaga S.,  
 RT "A novel protein toxin from the deadly box jellyfish (Sea Wasp,  
 Habu-kurage) Chirospalmus quadrangatus.";  
 RL Biosci. Biotechnol. Biochem. 66:97-102(2002).  
 CC -1- FUNCTION: Has potent hemolytic activity. lethally toxic to  
 CC crayfish when administered via intraperitoneal injection (LD50 =  
 80 microg/kg). Causes cutaneous inflammation in humans.  
 CC -1- SUBCELLULAR LOCATION: Cnidocyst and then secreted.  
 CC -1- PTM: There are disulfide bonds (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE JELLYFISH TOXIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB045319; BAB82520.1;  
 KW Toxin; Hemolysis; Signal.  
 FT SIGNAL 1 19  
 CHAIN 19 462  
 SEQUENCE 462 AA: 51604 MW: 461253DFCEB3D3C0 CRC64:

Query Match 42.7%; Score 992.5; DB 1; Length 463;  
 Best Local Similarity 43.5%; Pred. No. 2.2e-61;  
 Matches 197; Conservative 95; Mismatches 148; Indels 13; Gaps 7;

GenCore version 5.1.4-P5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2003, 11:14:48 ; Search time 25 Seconds

(without alignments)  
746.574 Million cell updates/sec

Title: US-09-647-522-5

Perfect score: 2322  
Sequence: 1 MILKHLPMFLFYLAITSKHK.....SGYRSMENNDPGGHSWIT 450

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2322	100.0	450	1	Q9GV72 carybdea ra
2	992.5	42.7	463	1	CYTX_CARLA
3	423.5	18.2	462	1	CYTX_CARLA
4	116.5	5.0	376	1	MDRC_YEAST
5	112.5	4.8	964	1	PMPE_CHLTR
6	110	4.7	1073	1	ACAG_ARATH
7	108.5	4.7	1023	1	HLXA_ECOLI
8	108.5	4.6	1024	1	HLXA_ECOLI
9	106	4.5	1369	1	MSH5_CAEEL
10	105.5	4.5	312	1	PAS3_RHOFA
11	105	4.5	506	1	ABR_ECOLI
12	103	4.4	1456	1	RPO_PVACP
13	102	4.4	1456	1	RPO_PVACP
14	101.5	4.4	299	1	PYRB_ARCIB
15	101	4.3	393	1	FLAA_PSEAE
16	101	4.3	535	1	HRLI_HAANI
17	101	4.3	1520	1	ACPD_VIBCH
18	100.5	4.3	909	1	SBCC_DEIRA
19	100	4.3	437	1	YABE_BACSU
20	100	4.3	444	1	Y808_CHLPP
21	99	4.3	948	1	PMA5_ARATH
22	99	4.3	1935	1	MYSS_CYPCA
23	98.5	4.2	379	1	FLAA_VIBCH
24	97.5	4.2	487	1	SR54_ENTHI
25	97	4.2	825	1	GNH7_BACSA
26	97	4.2	1934	1	MYH7_MESAU
27	97	4.2	1935	1	MYH7_RAT
28	96.5	4.2	1065	1	NOHG_RHIME
29	95.5	4.1	3027	1	POG6_PYPV1
30	95	4.1	381	1	YBD2_YEAST
31	95	4.1	434	1	T147_HUMAN
32	95	4.1	483	1	KICL_MOUSE
33	95	4.1	675	1	ATKB_DEIRA

## ALIGNMENTS

RESULT 1	ID	CTX1_CARLA	STANDARD:	PRT:	450 AA.
AC	Q9GV72	CTX1_CARLA			
DT	15-JUN-2002 (Rel. 41, Created)				
DT	15-JUN-2002 (Rel. 41, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Toxin-1 precursor (CTX-A / CTX-B).				
GN	CR1-1				
OS	Carybdea rastoni (Box jellyfish).				
OC	Eukaryota; Metazoa; Cnidaria; Cubozoa; Cubomedusae; Carybdeidae;				
OC	Carybdea.				
OX	NCBI_TaxID=78582;				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	TISSUE=Cnidoblast;				
RX	MEDLINE=20422301; PubMed=10964707;				
RA	Nagai H., Takawa K., Nakao M., Ito E., Miyake M., Noda M., Nakajima T.;				
RT	"Novel proteolase toxins from the box jellyfish (sea wasp) Carybdea rastoni".				
RL	Biochem. Biophys. Res. Commun. 275:582-588(2000).				
CC	-1- FUNCTION: Has potent hemolytic activity. lethally toxic to mice (LD50 = 20 microg/kg via i.v.) and grayfish (LD50 = 5 microg/kg via intraperitoneal injection). Causes cutaneous inflammation in humans.				
CC	-1- SUBCELLULAR LOCATION: Cnidocyst and then secreted.				
CC	-1- TISSUE SPECIFICITY: Synthesized in the tentacle then migrates to the cnidocyst.				
CC	-1- PTM: There are disulfide bonds.				
CC	-1- SIMILARITY: BELONGS TO THE JELLYFISH TOXIN FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	EMBL: AB015878; BAB12728.1; -				
DR	TOXIN: Hemolysis; Signal.				
KW	SIGNAL				
FT	CHAIN 1 450				
FT	TOXIN-1.				
SO	SEQUENCE 450 AA; 49392 MW; CD393CF25BFD2FD CRC64;				
QY	Query Match	100.0%	Score 2322;	DB 1;	Length 450;
QY	Best Local Similarity	100.0%	Pred. No. 2.8e-153;		
QY	Matches 450; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
DB	1 MILKHLPMFLFYLAITSKHKGRSDVNSILITVERALKEASSNDALEALGELGEIOT 60				
QY	61 KDRVGOATKILGVSALGKLSNGDATRIISGCDIYAGIATPFGPGVMGIGAVASFV 120				



Query Match 4.88; Score 112.5; DB 2; Length 964;

```

Query Match          4.8%   Score 112.5; DB 2; Length 6713;
Best Local Similarity 21.7%; Pred. No. 1.le+02;
Matches    92; Conservative    55; Mismatches 157; Indels 119; Gaps 17.;
QY      16 TSAAKGRSDVNSLTFTKETALKEKSGN-----EALF-----ALEG----- 53
       |  | : : : | : : : : ||| | | |
Db      2037 TEARDKOTATNDAYTAAKTLLDTAGSNDNRKAVEAQLORNTAKTALNGDARLEAKN 2096
               -----PDRVGOAKRIILGSVG--SALGKLIN-----SGDAT 88
QY      54 -LKGEIQT-----
Db      2097 TAKOOLATMSHLTNAAQRANTLFEQIERGGTVAGYGICIANASTLIQNMAMQLOPSIASKAPT 2156

```

A:Gene: TM1143  
C:Superfamily: methyl-accepting chemotaxis protein

Query Match 5.0%; Score 116; DB 2; Length 530;  
Best Local Similarity 19.4%; Pred. No. 2.1;  
Matches 54; Conservative 63; Mismatches 122; Indels 40; Gaps 9;

QY 15 ITSAKHGRSDVNSLTFTVET---ALKEASGSENALEBLEKKEIQTKPRVQAKRI 71  
DB 248 ITNOLGISTKEMDNISTRIESISASVOETTAGSEISATKNTIADSAQOASPADOSTOL 307  
QY 72 LGSVSAIGKINSGDATGICGLDIVAGIATFGPGVGMGIGAVAFSSTLSLTGSS 131  
DB 308 AKEDADALKTYL--EYTRKISNSAKDVERVVSF---OKGBEITTSFETITNAT----- 356  
QY 132 AKNSVAAYIDRALSKHRDEAIOHRAAGAKRDESSAFIOVKQOOSNLTDSDLSITANV 191  
DB 357 AEQTLLALNMAIEAR-----AGEAGRGFAVVADEIRKLAESQASENVRVYNEI 409  
QY 192 ----PVYKFSNFIOLESRIISOGAATTSLSDAK-----RAVDFTLLYCOLVYMR 238  
DB 410 RSIADAGKVS---EITARVEEGTKLADDEDEKINSIVGAVERINEMQNTIAALEEQT 466  
QY 239 LLYD---LAILYRKGNAEHVASAVENANRKEIADT 273  
DB 467 AAVDEITTRMTENAKNAEETINSKEVNAKRLQEISAST 504

## RESULT 10

AC1647  
Lactobacillus phage phi-gle minor capsid protein 1608 homolog 11n1716 [Imported] - Liste

C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AC1647  
R:Glasser, P.; Fraigneul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bjoeker  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.  
Science 294, 849-852, 2001  
A:Authors: Krett, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Me  
ok, C.; Schlueter, T.; Simoes, N.; Tilleret, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlund,  
A>Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AC1647  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1571 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC96947.1; PID:g16414203; GSPDB:GN00178  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: 11n1716

Query Match 5.0%; Score 116; DB 2; Length 1571;  
Best Local Similarity 20.3%; Pred. No. 8.9;  
Matches 84; Conservative 74; Mismatches 148; Indels 108; Gaps 18;

QY 11 IVLAITSANHGKRSVDNSLTFTVETALKEASGSENALEBLEKKEIQTKPRVG--QA 68  
DB 141 LVVAGKRAFENYASDTNESINKVDVAFKNA-----ESVKKWSKTTLLNIGLAQG 190  
QY 69 TKI-----LGSVSAIGKINSGDATKIIISGLDIVAGIATFGPGVGMGIGAVAFS--- 120  
DB 191 TALDLAATYGDMSTSMG--ISTOEAKMSTSMVLA-----GNLAKSRNID 234  
QY 121 ----SSILSLFTG--SSAKNSVAVIDR-----ALSKHRDEAIOHRAAGAKRDESSA 168  
DB 235 IDBANALINGVFTGETEALAKSLGIVMTQTLEQFALETGAGKVAKSSTEYTKONIAE-- 292  
QY 169 FIOVMQOOSNLTDSDLSITANVPYKFSNFIOLESRIISOGAATTSLSAKKRAVDFTLL 228  
DB 293 ----KAOKLNEAIRKEHGKNSLEAREAKOIKOIEIAKTSIG--AKVNLKMKO--DELVR 344  
QY 229 YCOLVYMRFTLVDLAILYRKGNAEHVASAVENANRV---NRELADTLDFTLHKLIPEQ 284

DB 345 LRYNYVMKOT-----TNAHDEFKNTSDQAANARVFSSESVELASNAGOFLLPII--T 395  
QY 285 ALICAVYHPISASENSKAILMTYTKYFGVDPVPRPIGNRRYFTNSYMTYSGSEAYM-- 342  
DB 396 PLI-----IQAISDETKKL-----SDIPGAVKGMKEKFKPWF--EVFTVGDFFPKD 439  
QY 343 -----GNMYFRCNSVNRPNIRNKRMSKSDGFYTMENSRRK 377  
DB 440 LIPSAKELAKSMGPGFIEGVLAFKGLDIVLNTVTYIPAFKAVYKTRRENDPSMK 493

## RESULT 11

AG0438  
Probable Rns accessory genetic element YP03606 [Imported] - Yersinia pestis (strain C

C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
C:Accession: AG0438  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tilball, R.W.; Holden, M.T.G.; Prentice, M  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G  
ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel  
Nature 413, 523-527, 2001  
A>Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AG0438  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-800 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC93075.1; PID:g15981527; GSPDB:GN00175  
C:Genetics:  
A:Gene: YP03606

Query Match 5.0%; Score 115; DB 2; Length 800;  
Best Local Similarity 17.3%; Pred. No. 4.2;  
Matches 101; Conservative 79; Mismatches 202; Indels 202; Gaps 24;

QY 12 VLAITSAKHGRSDVNSLTFTVETALKEA---SGSNEAL-----EALGKKEIQTKP 62  
DB 1 MFAHDKANNKNAADKTGVTPTPHIVSPEAPYVSGAOMTEAGIGDAVGTAAG----- 55  
QY 63 DRVQATKTLGVSGLKGLNSGDAT-----KIISGC--LDIVAGIATFGPGV 110  
DB 56 -RINAAEYAKTALDAKAVLDGGVTPINIVATGPAVTVSGAOMTEAGIGDAVGTAAG 114  
QY 111 MGIGAVAFSSTLSLTFTSSAKNSVAAYIDRALSKHRDEAIOHRAAGAKRDESSAF 170  
DB 115 KRINQAAE-----VAKTLEAKAKVLDGCVTPMN---LAAGGAGAGLPDSAAATIS 162  
QY 171 QVAKQOOSNLTDSDLSITANVPYKFSNFIOLESRIISOG--AATYSLDAKRAVDFTLL 229  
DB 163 RLVKOPSGL---QFTLTASLPQTFPAVVDFTLSEKLSPPVLNVLGASADPAVDPAV 219  
QY 230 ---COLVYMR-----TLVDLAILYRK 250  
DB 220 DEDATLFIWREGVLRSITGVNAVSFEQDGTGFHOTRYSMVIRPALMRTSLRRNARIF 279  
QY 251 NAEHVASAVENANRVK-----ELAADLDFLHKLIPEQAL----- 286  
DB 280 SVEEITTLKEGINDFAFGFRRHPVREFCVOIESPDTLQRLTAEGIFYYFEESA 339  
QY 287 -----IGAV-----YHP--ISASETSKAILMTYTKYGVDPVPRPIGNRRYFTN 328  
DB 340 GKNTVYADVGLPKGASLPPNPNVAQAQOELSITTFRRSAQVR--PAMVOLKDYTFKN 397  
QY 329 STW--NTYSCS-----EAYMGNYFRCNSVNRPNIRNKRMSKSDGF 367  
DB 398 PMAAAAFSEQSGELQNOQRPDYEHFDPGRKKAQOHQODEFTFRRLDLRNDANLGOASND 457  
QY 368 YTMENSDBRKLYTTKHDG--WGM-----GTLDEDPGDCG-----HMFIFPL 407  
DB 458 FTLOPQOLSTLY--NIPRODLNNAWOLLIQHSQKQMALEQASGQGVTLFHFHFFIFP- 514  
QY 408 RHGKYVSSKRWPNWMTWYESSASGYIRSWENNP-----GPO 444



Matches 100; Conservative 54; Mismatches 127; Indels 174; Gaps 23;

4 KHL-----PWLFTVLAITSAGHGRSDVNSLITKVTETALKEAG-SNEALEALEGLK---- 55  
 Db 225 KHLGVAPM-----DVVGGSGPESISLPSPSEPLSDLSKSTEAAPVCKRSIKKRV 286  
 QY 56 ---GEIQKPDVPGATKILGVSGLNSGDATKIISGCLDIYAGIATTFGGPVGNG 112  
 Db 287 FRSQLEPIPD-----SLDSEETISERLD-----STHHGAVGAG 321  
 QY 113 -----IGAVASVSSILSLFTGSSAKNSVAAYDR----- 142  
 Db 322 BCESTKESHSHIKKKKKKKKKKSVLATSSDSAVTDSKKN---ALVDSSESGAVREED 378  
 QY 143 -----ALSKHDEAIQR-----HAAGAKRDESSAFIOVMQOSNLTDSDLS 185  
 Db 379 VDHRAEAEAGACSTEKHR-EMQRLPEPTHEESNSSESNAARHISDRSDSDVD 437  
 QY 186 IIAANVPYKFSNFTGLESRISSGAATTSLSDAKRAVDLILYQQLVYMRFTLLVDLAI 245  
 Db 438 LGSA---VROLREFIPDIOER---AAT---IRMYRDLG- 469  
 QY 246 LYRGNMAEHA-----SAVENANRVNKEIADTLDLHLKILPEQALIGAVHPISASST 299  
 Db 470 LFEKKAQGVAIREFKFSNKE-----NKOIEKNVODFL-----SLTGIES 509  
 QY 300 SKAILNTYKTYGVDPVPRPIGN--RRYKFTNSYMTYSICSEAYMGVNYFRGCSNVNPN 357  
 Db 510 ADKLITYDRY---PEKTLITLTKRKHAFL-----RLHIGKIGIARWKVLY---Y 551  
 QY 358 IRVSKMSD-----GYTMENSRRKRLITYTKHDGM 387  
 Db 552 YRAKIFDVNNYKGRYNEEDTKKLKAYHSLHGNMW 586

RESULT 8  
 DEBYWC  
 malate dehydrogenase (EC 1.1.1.37), cytosolic - yeast (Saccharomyces cerevisiae)  
 N/Alternate names: protein AOE423; protein 00537; protein YOL126c  
 C/Species: Saccharomyces cerevisiae  
 C/Date: 30-Jun-1993 #sequence\_revision 19-Jul-1996 #text\_change 16-Jun-2000  
 C/Accession: S63444; S12937; A34986; S05770; S66823; S71982  
 R/Casamayor, A.; Khalid, H.; Balcells, L.; Aldea, M.; Casas, C.; Herrero, E.; Arino, J.  
 submitted to the EMBL Data Library, November 1995  
 A/Description: Sequencing of a 13.4 kbp fragment of the left arm of chromosome XV reveal  
 A/Reference number: S63440  
 A/Molecule type: DNA  
 A/Accession: S63444  
 A/Residues: 1-423 <CAS>  
 A/Cross-references: EMBL:U41293; NID:g1209710; PIDN:ACC49466.1; PID:g1209715  
 A/Experimental source: strain FY1679  
 R/Minard, K.I.; McAlister-Henn, L.  
 Mol. Cell. Biol. 11, 370-380, 1991  
 A/Title: Isolation, nucleotide sequence analysis, and disruption of the MDH2 gene from S.  
 A/Reference number: S12937; MUID:91094852; PMID:1986231  
 A/Accession: S12937  
 A/Molecule type: DNA  
 A/Residues: 47-423 <MINI>  
 A/Cross-references: EMBL:M62808; NID:g171915; PIDN:AAA34766.1; PID:g171916  
 A/Accession: A34986  
 A/Molecule type: protein  
 A/Residues: 48-73 <MIN2>  
 R/Bopetzki, E.; Entian, K.D.; Lotspeich, F.; Mecke, D.  
 Biochim. Biophys. Acta 912, 398-403, 1987  
 A/Title: Purification procedure and N-terminal amino acid sequence of yeast malate dehyd  
 A/Reference number: S05770; MUID:87185517; PMID:3552052  
 A/Accession: S05770  
 A/Molecule type: protein  
 A/Residues: 48-73, 'P', 75-81 <KOP>  
 R/Arino, J.; Casamayor, A.; Gamo, F.J.; Gancedo, C.; Lafuente, M.J.; Aldea, M.; Casas, C.  
 submitted to the Protein Sequence Database, July 1996  
 A/Reference number: S66814  
 A/Accession: S66823

A/Molecule type: DNA  
 A/Residues: 1-423 <ARI>  
 A/Cross-references: EMBL:274868; NID:g1420008; PIDN:CA99145.1; PID:g1420009; GSPDB  
 A/Experimental source: strain S288C  
 R/Casamayor, A.; Khalid, H.; Balcells, L.; Aldea, M.; Casas, C.; Herrero, E.; Arino,  
 Yeast 12, 1013-1020, 1996  
 A/Title: Sequence analysis of a 13.4 kbp fragment from the left arm of chromosome X  
 ending frames.  
 A/Reference number: S71978; MUID:97051588; PMID:8896265  
 A/Accession: S71982  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-423 <CAW>  
 A/Cross-references: EMBL:U41293; NID:g1209710; PIDN:ACC49466.1; PID:g1209715  
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, November 199  
 C/Genetics:  
 A/Genes: SGD:MDH2; MIPS:YOL126c  
 A/Cross-references: SGD:S0005486; MIPS:YOL126c  
 A/Map position: 15L  
 C/Suprafamily: L-lactate dehydrogenase  
 C/Keywords: cytosol; homodimer; NAD: oxidoreductase  
 F:48-423/Product: malate dehydrogenase #status experimental <MAT>  
 F:228,201/Active site: Asp, His #status predicted  
 F:231/Binding site: substrate (Arg) #status predicted

Query Match 5.0%; Score 116.5; DB 1; Length 423;  
 Best Local Similarity 20.2%; Pred. No. 1.4;  
 Matches 72; Conservative 65; Mismatches 125; Indels 95; Gaps 17;

QY 13 LAITSAGHGRSDVNSLITKVTETALKEAGS-----NEALEALEGLGEIOTRPD 63  
 Db 62 IAILGAAGIGQSLILKLAQLOLYKRESNRSVTHILALVDVNGEALNGVTADLSHDT 121  
 QY 64 RVGQATKILGVSGLKLNCGDATKISG-----CLDIYAGIATTFGGPVGNG 112  
 Db 122 PIVSSH--SPAGGIENCLHNASIVYIPAGVRRKRGMTRDDLFVNAGIISQLDSTI--- 176  
 QY 113 IGAVASVSSILSLFTGSSAKNSVAAYDRALSKH---RDEAIQRHAAG-AKRFPAESSA 168  
 Db 177 --ACCDLSKVFVLYISNPVNSVLPVWVNSLTKHNPGRNGIERIRNGVTKLDIVRST 234  
 QY 169 FIQVMQOSNLTD-----SDLSITAN-----VPYKFSNFTGOL----- 203  
 Db 235 FLREINIESGLTPRNSMPDVPVIGSHGETIIPFSQSNFLRNEDQLYLHRYQYG 294  
 QY 204 -----ESRISGGAFTSLSDA--KRAVDLILYQQLVYMRFTLLVDLAILRKNAEHA 256  
 Db 295 GDEVYKANKGSGATLSMAHAGYCVQFV-----SLL-----GNIQIH 335  
 QY 257 SA---VENANRVNKEIADTLDLHLKILPEQALIGAVHPISASSTSKAILNTYKY 309  
 Db 336 GTTYVPLKDNANPFIARGAD-----QLLP--LVDGADYFAIPPLITTKGV-STVDY 383

RESULT 9  
 C72291  
 methyl-accepting chemotaxis protein - Thermotoga maritima (strain MSB8)  
 C/Species: Thermotoga maritima  
 C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
 C/Accession: C72291  
 R/Nelson, K.E.; Clayton, R.A.; Gail, S.R.; Gwin, M.L.; Dodson, R.J.; Haf, D.H.; H  
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
 Nature 399, 323-329, 1999  
 A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
 A/Reference number: A72200; MUID:99287316; PMID:10360571  
 A/Accession: C72291  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-530 <ARN>  
 A/Cross-references: GB:AE001771; GB:AE000512; NID:g4981678; PIDN:AAD36219.1; PID:g49  
 A/Experimental source: strain MSB8  
 C/Genetics:





QY 421 NMFMMESSASGYIRSWENPPOGHWST 450  
 DB 421 NMFMMESSASGYIRSWENPPOGHWST 450

## RESULT 2

toxin-A - jellyfish (Carybdea alata)

C:Species: Carybdea alata  
 C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 17-Nov-2000  
 C:Accession: J07372; PC7095  
 R:Nagai, H.; Takawa, K.; Nakao, M.; Sakamoto, B.; Crow, G.L.; Nakajima, T.  
 Biochem. Biophys. Res. Commun. 275, 589-594, 2000  
 A:Title: Isolation and characterization of a novel protein toxin from the Hawaiian box  
 A:Reference number: J07372  
 A:Accession: J07372  
 A:Molecule type: mRNA  
 A:Residues: 1-463 <NAG>  
 A:Cross-references: DDBJ:AB036714  
 A:Accession: PC7095

A:Molecule type: protein  
 A:Residues: 158-168:431-463 <NAG>  
 C:Comment: This protein, belongs to a class of bioactive proteins with hemolytic activity  
 C:Keywords: hemolysis; inflammation; toxin

Query Match 42.7%; Score 992.5; DB 2; Length 463;  
 Best Local Similarity 43.5%; Pred. No. 76-62;  
 Matches 197; Conservative 95; Mismatches 148; Indels 13; Gaps 7;

QY 9 LRTVLAITAKHG-----KRDVNSLITVEYALAKAS--GSNEAL-ELALELKEIQT 60  
 DB 11 LFLVISTAPPSQARLSRYRRSADAVSTIDIGIQLNDLGTPTKRLKEALQGVQAVRK 70  
 QY 61 KP-DRVGQATKILGVSALGKLSGDAITKISGCDIYAGIATTEGPGVAGKIGAVAF 119  
 DB 71 EATITAKSTIVSGSLSKFSQDPDVASGCDITIASVATYGGPGIAGIAGVASTL 130  
 QY 120 VSSITSLFTGSSAKNSVAVIDRALSKRDEAIQRHAGAKRDPFASSAFIOYMKQOONL 179  
 DB 131 ISSITSLFSGNSMGSAIKVIDDAFKRYRQDELDVAKAKRFENAVITFVNSVSTENL 190  
 QY 180 TPSSDIITANVPYKFSNFIFIGOLESRISOGAATSTLSAKRANVDILLYCOLVVRKEL 239  
 DB 191 TEVHLDSVADAVPDAFTMLGLVESRLNRSVSTDNNAKPTINFIFLYLDSVARETL 250  
 QY 240 LVDLAILYRK--GNAEHVASAVENANRVKELADTLDLHLKILPEQALIGAVYHPISAS 297  
 DB 251 LQVILLKVRAGAVDELALSLTSDONKEATREIVTFTHQMETKYSILCGSYTYPTDHS 310  
 QY 298 ETSKAILNTKTFYGPDPVPR-PIGNRRYKFTNSYNTYSICSEAYMGNYMFRGCSNVRNP 356  
 DB 311 KAAIILKILTKFFGVDPDARYTFDGLLYMQNRANRYSICKESYAGNMFRGCKDSSYH 370  
 QY 357 NRVKMSDGEFTYMTNSDRKLYITRKHOQMGWGLTDEPQGGQHHKRFPLRHGKYMSS 416  
 DB 371 GIRIKLNGEYITIT-TLSKAMTYVTKHOGMGWGTADDEPGBGYETFLPLNGEYIMST 429  
 QY 417 KRPMWFMYMESSASGYIRSWENPPOGHWMSI 449  
 DB 430 KKMFDYFYMESSAHGYIRSWHYNPDPOGOWKI 462

## RESULT 3

toxin-A - Chitropsalmus quadrifidatus

C:Species: Chitropsalmus quadrifidatus  
 C:Date: 03-Jun-2002 #sequence\_revision 03-Jun-2002 #text\_change 03-Jun-2002

C:Accession: J07805; PC7181  
 R:Nagai, H.; Takawa, Kuroda, K.; Nakao, M.; Oshiro, N.; Iwanaga, S.; Nakajima, T.  
 Biosci. Biotechnol. Biochem. 66, 97-102, 2002  
 A:Title: A novel protein toxin from the deadly box jellyfish (see wasp, habu-kurage)

A:Reference number: J07805; PMID:1186126; MUID:21554932  
 A:Accession: J07805

A:Molecule type: mRNA  
 A:Residues: 1-462 <NAG>  
 A:Cross-references: DDBJ:AB045319  
 A:Accession: PC7181

A:Molecule type: protein  
 A:Residues: 432-452:21-54 <NAG>  
 C:Comment: This protein is a major proteinaceous toxin in the nematocyst, so that 1  
 C:Genetics:  
 A:Gene: Cqtx-A  
 F:1-20/Domain: signal sequence #status predicted <SIG>  
 F:45-62/Region: amphiphilic alpha-helix-like sequence #status predicted

Query Match 18.2%; Score 423.5; DB 2; Length 462;  
 Best Local Similarity 26.3%; Pred. No. 56-22;  
 Matches 120; Conservative 108; Mismatches 171; Indels 57; Gaps 17;

QY 24 SDVNSLITVEYALKEASGSENEALALKEIOTKPDVQATKILSGVS---ALG 80  
 DB 26 SGLASLAKIDA--KRPSG--KQLFQVANNMQKIEKFSNDRKAKVMKALISLSTAVG 81  
 QY 81 KLSGDPATKIIISGCDIYAGIATTEGPGVAGKIGAVAFVSSILSLFTGSSAKNSVAAYI 140  
 DB 82 KQSGDPARIASGCDIILVIGISSVLKD--FAKSPIFSLISVYGLFSGTKAESVGSVY 139  
 QY 141 DRALSKRDEAIQRHAGAKRDPFASSAFIO--VMKQOONLTSDSITIANVPYKFSNF 199  
 DB 140 KRYVQEOSQOEALGYKREYVASKATLDGRNETSDLSPEVSALGANVPYGVRF 199  
 QY 200 IOLESRIISOGAATSTLSAKRANVDILLYCOLVVRKELVLDL-AILYRKNAEHVASA 258  
 DB 200 IAWVQRIKRRPRTE-SEIKRVLSMLEFTDLCSIRDILIDVQLVATPCHSPRIASG 258  
 QY 259 VENANRVKELADTLDLHLKILPEQALIGAVY-HPISASETSKALNTKYKGVDPVPR 317  
 DB 259 IKEYSNLGRREYKVFEDLKTNDKETYLELSLYLPRERNEOSORIF---DLAK 312  
 QY 318 PIGNRRYK-----FTNSYNTYSICSEAYMGNYMFRGCSNVRNPNTRVKMSDGEFT 369  
 DB 313 VAYDRLKODLTGIVFSSLHMPNFTLCSK---DYALITCKPYG-SLRLOKLDNGFSS 368  
 QY 370 MENSRR-----KLYTKHDQMGWGLTDEPQGGQHHKRFPLRHGKYMSS 418  
 DB 369 IKTQSNPRVCHRYGEYILFTHNRN-----DLEKFNFPVKLGERKIYLLSSRA 418  
 QY 419 WPMFMYMESSASGYIRSWENP-----GPOGHWMSI 449  
 DB 419 SPNKFAIVPKTAKGDLFFVYDGIPIQLGYGNOGYFTL 454

## RESULT 4

hypothetical protein PX01-79 - Bacillus anthracis virulence plasmid PX01

C:Species: Bacillus anthracis  
 C:Date: 12-Nov-1999 #sequence\_revision 12-Nov-1999 #text\_change 16-Feb-2001

C:Accession: G59100  
 R:Okunaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Kelm, P.; Koe  
 J. Bacteriol. 181, 6509-6515, 1999  
 A:Title: Sequence and organization of PX01, the large Bacillus anthracis plasmid har  
 A:Reference number: A59091; MUID:99445483; PMID:10515943  
 A:Accession: G59100

A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1222 <OKI>  
 A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD32383.1; PID:g4894295  
 A:Experimental source: strain Sterne  
 A:Note: similar to hypothetical, hydrophobic protein (567 aa), Bacillus firmus (U645  
 C:Genetics:  
 A:Gene: PX01-79  
 A:Genome: plasmid  
 C:Superfamily: hypothetical protein PX01-79

Query Match 5.98%; Score 137; DB 2; Length 1222;  
 Best Local Similarity 23.0%; Pred. No. 0.21;

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2003, 11:14:48 ; Search time 21 Seconds  
(without alignments)  
2060.023 Million cell updates/sec

Title: US-09-647-522-5

Perfect score: 2322

Sequence: 1 MILKHPMLFIVLAITSRAK.....SGYIRSWNNPQGHWSIT 450

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR-73:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2317	99.8	450	2 JC7371	toxin - jellyfish
2	992.5	42.7	463	2 JC7372	toxin-A - jellyfish
3	423.5	18.2	462	2 JC7805	toxin-A - Chitropsa
4	137	5.9	1222	2 G59100	hypothetical prote
5	119.5	5.1	595	2 AE0150	probable methyl-ac
6	118.5	5.1	506	2 AG0893	acetylcholinesterase
7	117	5.0	833	2 S54776	transcription factor
8	116.5	5.0	423	1 DEBYC	malate dehydrogenase
9	116	5.0	530	2 C72291	lactobacillus phag
10	116	5.0	1571	2 AC1647	methyl-accepting c
11	115	5.0	800	2 AG0438	probable Rhs acces
12	112.5	4.8	510	2 C84738	probable kinesin I
13	112.5	4.8	964	2 E71460	probable outer mem
14	112.5	4.8	6713	2 B89921	hypothetical prote
15	111.5	4.8	656	2 A72428	methyl-accepting c
16	111.5	4.8	1336	2 T18288	ABC transporter
17	110.5	4.8	656	2 E72379	methyl-accepting c
18	110.5	4.8	661	2 G72316	methyl-accepting c
19	110.5	4.8	1024	2 S10056	hemolysin A - Esch
20	110	4.7	810	2 B84185	cytochrome-like pr
21	110	4.7	2297	2 AB2494	hypothetical prote
22	109	4.7	541	2 E83032	probable chemotaxi
23	109	4.7	612	2 B75282	probable acyl-CoA
24	108.5	4.7	659	2 G82365	methyl-accepting c
25	108.5	4.7	1023	1 LBECA	hemolysin A - Esch
26	108	4.7	492	2 A97429	flag protein prote
27	108	4.7	492	2 AB2647	hook associated pr
28	107.5	4.6	2155	2 C97523	hypothetical prote
29	107.5	4.6	2155	2 AD2742	conserved hypothet

30	106.5	4.6	669	2 AC2624	methyl-accepting c
31	106.5	4.6	692	2 A97406	hypothetical prote
32	106.5	4.6	1098	2 B70232	hemagglutinin/hemo
33	106	4.6	1975	2 B81192	hypothetical prote
34	105.5	4.5	312	2 C55578	aerotaxis receptor
35	105	4.5	506	2 E65095	aerotaxis sensor r
36	105	4.5	506	2 A85968	aerotaxis sensor r
37	105	4.5	506	2 B91123	ABC transporter
38	104.5	4.5	532	2 E87343	methyl-accepting c
39	103	4.4	504	2 H87422	hypothetical prote
40	103	4.4	1456	2 S14005	leucine dehydrogen
41	102.5	4.4	353	2 A11859	hypothetical prote
42	102.5	4.4	460	2 T19111	flagellin, 40K - P
43	102	4.4	394	2 A37853	hypothetical 16S.1
44	102	4.4	1456	2 JQ2284	aspartate carboxyl
45	101.5	4.4	299	2 B69263	

## ALIGNMENTS

## RESULT 1

toxin - jellyfish (Carybdea rastoni)

C:Species: Carybdea rastoni

C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 17-Nov-2000

C:Accession: JC7371; PC7094

R:Nagai, H.; Takawa, K.; Nakao, M.; Ito, E.; Miyake, M.; Noda, M.; Nakajima, T.

Biochem. Biophys. Res. Commun. 275, 582-588, 2000

A:Title: Novel proteinaceous toxins from the box jellyfish (sea wasp) Carybdea rast

A:Reference number: JC7371

A:Accession: JC7371

A:Molecule type: mRNA

A:Residues: 1-450 <NAG>

A:Cross-references: DDBJ:AB015878

A:Accession: PC7094

A:Molecule type: protein

A:Residues: 39-55;56-70;196-210;250-267;268-279;309-325;363-377;378-382 <NAG>

C:Comment: This protein, a member of bioactive protein, has hemolytic activity.

C:Keywords: hemolysis; inflammation; toxin

Query Match 99.8%; Score 2317; DB 2; Length 450;

Best local similarity 99.8%; Pred. No. 1.2e-154;

Matches 449; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MILKHPMLFIVLAITSRAKGRSDVNSLITRETFALKEAGSNEALALEGLGEIOT	60
DB	1	MILKHPMLFIVLAITSRAKGRSDVNSLITRETFALKEAGSNEALALEGLGEIOT	60
QY	61	KPDPRGQATKILGVSALGKLNISGDAIKIISGCDIYAGIATTFGGPVGIGAVASPV	120
DB	61	KPDPRGQATKILGVSALGKLNISGDAIKIISGCDIYAGIATTFGGPVGIGAVASPV	120
QY	121	SSILSFTGSSAKNSVAVIDRALSKHRDEALIORHAAKRDPAESSAFIOYMOOSNLT	180
DB	121	SSILSFTGSSAKNSVAVIDRALSKHRDEALIORHAAKRDPAESSAFIOYMOOSNLT	180
QY	181	DSDSLIIANVPYKFSNFICQLESRIISGAAITSLSAKRAVDITLLYCOLVYVRETL	240
DB	181	DSDSLIIANVPYKFSNFICQLESRIISGAAITSLSAKRAVDITLLYCOLVYVRETL	240
QY	241	VDLALILYRKGAHEVASAVENANRNKELADTDLFLKLLPEOLIGAVYHPIASSETS	300
DB	241	VDLALILYRKGAHEVASAVENANRNKELADTDLFLKLLPEOLIGAVYHPIASSETS	300
QY	301	KAILNYYTFYGVDPDRIGNRKRYKFTNSYNTYSICSEAYMGVYMGCSNVNPNIRV	360
DB	301	KAILNYYTFYGVDPDRIGNRKRYKFTNSYNTYSICSEAYMGVYMGCSNVNPNIRV	360
QY	361	SKMSGCEYTMESDRRKLYITRKHOOGMGCTLDEPQGGHRRFPLHGGYVMSRRMP	420
DB	361	SKMSGCEYTMESDRRKLYITRKHOOGMGCTLDEPQGGHRRFPLHGGYVMSRRMP	420

Best Local Similarity 21.9%; Pred. No. 39;  
Matches 75; Conservative 40; Mismatches 127; Indels 100; Gaps 16;

```

QY 42 GSENALEALEGKKEIQTKPDR---VGQATKILGSGSALG--KL--NSGDATKI---I 91
Db 70 GISEA-----DGGKGGANARCKSIAGIDIAQALGSGSIAIGDNKIYHNSNNNANIGAKA 124
QY 92 SGCLDIYAGIATTFGGPVMGIGA-----VAFVSSILSLFTGSSAKNSVAAYIDRA 143
Db 125 SGNESIAIGGDVLAGSHASIAIGSDDLTKKETVOQISELPIIRGQKALNDIYQLADTN 184
QY 144 LSKHDEAIORHAGAKRDFAESAFIQVMKQOSNLTBSDLITIANVPYKESNFIGOL 203
Db 185 LQKYRRTHAQGHASTA-----VGAMSTAKGHFSNAFGTR 218
QY 204 ESRIQGAATTSLSDAKRAVDLILYCOLVVRRETLVLAITYRKNAEHV---ASAVE 260
Db 219 ATAECTYSLAVGLTATAKAASSIAV-----GSNAQAIGFAATYAG 258
QY 261 NANRYNKLADTLEFLKLIPEQALIGAVYHPISASETSKALINITYYFGVDPVPRPIG 320
Db 259 GSTOVNLRGI-ALGFGSQVLQKNDVNAA-----NVRAY--APDDNQPID 301
QY 321 NRRYK--FTNSYWNYYISCSSEAYMGNYMFRCGCSNVRNPNIYV 360
Db 302 N-RYKATFKNGATDVFSI-----GN--SNGNDSIRKRIINV 334

```

Search completed: May 19, 2003, 11:28:04  
Job time : 69 secs

```
RESULT 13
US-09-815-242-12611
; Sequence 12611, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELTRA-011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12611
; LENGTH: 3158
; TYPE: PR
; ORGANISM: Staphylococcus aureus
US-09-815-242-12611

Query Match
Best Local Similarity 4.3%; Score 100; DB 10; Length 3158;
Matches 66; Conservative 65; Mismatches 147; Indels 54; Gaps 14;

QY 17 SAKHGRSDVNSLTETKVFALKPSGNSNE-----ALEALEGKGIQTK-PDRVG 66
DB 1054 TGGNANQPEVQAITRYVOTTLQALGDHNLQVAKTNATQALDALTSLNDOKTALKDOVT 1113
QY 67 QATKILGVSAGLKL--NSGDATKTIISGCLDIVAGIATFGG-----PYGMIGAV 116
DB 1114 AATLV-----TAVHQLQONANTLNQAMHGLRESIODNAATKANSKIYNEDQPEQONTDOA 1168
QY 117 ASFVSSILSFTG---SSAKNSVAAYIDRA-LSKRHDEAIORHAAGAKRDFASSAFIOY 172
DB 1169 VQANSTIINQATLNDNNALNQAAATYNTKALHGDVYKQNDKAKQVQSOLAYLNNA 1228
QY 173 MK-QQSNLSDSLIITANPVYKFSNFICQLSRSISQGAATISLSAKAAVDFIL----- 227
DB 1229 QKHEMTLLDSEETTRIAVKODLTE-AQALDQUNNTLQOOSIAD---KDATPASSAYVAEP 1284
QY 228 ----LVCQAVMRETLVLVDL-ALLYRKNAEHVAASAVENTNR-----VNKELADT 273
DB 1285 NKQAVDEAVQNAESIITAGINPTINKGNVSSATQAVTTSKNGLDGVERLAQDKQTAGNS 1344
QY 274 LDFLHLKLP--EQALIGAVYHPISASETSKAI 303
DB 1345 LNHLDQLTFAQQOQALEMQINNATTPRKVAEII 1376

RESULT 14
US-10-125-692-13
; Sequence 13, Application US/10125692
; Publication No. US20030044429A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Agerem, Alan
; APPLICANT: Hayashi, Fumitaka
; APPLICANT: Smith, Kelly D.
; APPLICANT: Underhill, David M.
; APPLICANT: Ozinsky, Adrian
; TITLE OF INVENTION: Toll-Like Receptor 5 Ligands and Methods
; FILE REFERENCE: P-15 5155
; CURRENT APPLICATION NUMBER: US/10/125,692
; CURRENT FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: US 60/285,477
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 379
; TYPE: PR
; ORGANISM: V. cholerae
US-10-125-692-13

Query Match
Best Local Similarity 4.2%; Score 98.5; DB 9; Length 379;
Matches 66; Conservative 37; Mismatches 85; Indels 131; Gaps 12;

QY 13 LATTSAKHGRSDVNSLTETKVF-ETALKEASGNSNEAA-LEALEGKGIQTKPDRVQAT- 69
DB 72 ISTAQTBEGAMNSETIILQMRDLALOSANGTNSASEROLNBSVAALODELRIRIETTS 131
QY 70 ---KIL-GSVGSLAKLNSGDATKIISG----- 93
DB 132 FGGRKILNLSFGASFOIGSSGEALIMGLTSVRADFRMGQSFIAEQPKTKEMGVPT 191
QY 94 -----CLDIVA----- 99
DB 192 ARDLKEFTKDEAVYLDIARDGDDIEELATYINGQTDLFKASVDOEGKLIQFVAEPN 251
QY 100 -----GIATTFGGPYVMGIGAVASFVSSILSFTGSSAKNSVAAYIDRALSKRDE 150
DB 252 IEGNFNISGLATELGLNGSPGVKTYVODI-----ITSVGSONAVGITDAL-KYVDS 305
QY 151 AIORHAAGAKRDFESSAFIOVAKQOONLTDSDLSITIANVPYK-----FSNFIGOL 203
DB 306 --QRADLGAKQNLRSI-----SNLSINQENVEAKSRKIKDTPRAKETTLQ 350
QY 204 -ESRISOGAATTSLSDAKR 221
DB 351 TKSQILQOAGTSIIAQAKQ 369

RESULT 15
US-09-813-214A-9
; Sequence 9, Application US/09813214A
; Patent No. US20020177200A1
; GENERAL INFORMATION:
; APPLICANT: Tucker, Kenneth
; APPLICANT: Piosilla, Laura
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE PROTEIN-106 POLYPEPTIDE,
; FILE REFERENCE: 7969-089-999
; CURRENT APPLICATION NUMBER: US/09/813,214A
; CURRENT FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: 08/968,685
; PRIOR FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 9
; LENGTH: 2122
; TYPE: PR
; ORGANISM: Moraxella catarrhalis
US-09-813-214A-9

Query Match
Best Local Similarity 4.2%; Score 97; DB 9; Length 2122;
```

PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO: 5639  
LENGTH: 2086  
TYPE: PRN  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5639

Query Match 4.5%; Score 104.5; DB 10; Length 2086;  
Best Local Similarity 20.7%; Pred. No. 8.9;  
Matches 79; Conservative 59; Mismatches 134; Indels 109; Gaps 18;

QY 12 VLATSAKGRSDVNSLITKVTALKKASGN-----EALALEG-----L 54  
DB 1095 VINAASNPMDANAIQATVSTKRNALDGTNLTQAKOTATNAIDGATNLNKAQKDL 1154  
QY 55 KEIOTKPRVGOATKI---LGSVSGALGKLSNG---DATKISGLDIYAGIATTEGG 107  
DB 1155 KAOV-TSAORVANVTISQOTANELNLTAMGOLQHGIDENATKOTOKYRAEOSKRTAYD- 1212  
QY 108 PYGMCIGAVASFVSSILSLFTGSSAKNSVAAYIDRALSK-----HRDEAIQRHAGA 159  
DB 1213 ---QAVVA-AKAIILNKQTS---NSDKAAVDRALQOVSTKDALNDALAEKAA 1262  
QY 160 KRDF-----AESAF-----IOVMKOOSLTDSDLSITANV---PYVK 195  
DB 1263 KQNLGTLNHTNAORTALEGOINQATTVDGVTNKTANLTGAMNSLOGSINDKQATLR 1322  
QY 196 FSNFISGLESRI-----OGATISLDAKRAVDLILYCOLVARETL- 239  
DB 1323 NQNTDADESKRNATQAVTAAGLNLKQGTGNTSADVDNLTNTYRAKALNGAENLR 1382  
QY 240 -----LVDLAILYRKGNAEHVASAVENANRVN--KELADLDFLHKLIPQA 285  
DB 1383 NTKSATNTINGLPMLTOL-QKDNLKHQVEQAOVAVGVAVD-KGNTLN-----T 1431  
QY 286 LIGAVYHPIASSETSKALINY 306  
DB 1432 AMGALRTSIQNDNTTKTSQNT 1452

RESULT 11  
US-10-125-692-14  
Sequence 14, Application US/10125692  
Publication No. US20030044239A1  
GENERAL INFORMATION:  
APPLICANT: Aderem, Alan  
APPLICANT: Hayashi, Fumitaka  
APPLICANT: Smith, Kelly D.  
APPLICANT: Underhill, David M.  
TITLE OF INVENTION: Toll-Like Receptor 5 Ligands and Methods  
FILE OF INVENTION: of use  
FILE REFERENCE: P-1S 5155  
CURRENT APPLICATION NUMBER: US/10/125,692  
PRIOR FILING DATE: 2002-04-17  
PRIOR APPLICATION NUMBER: US 60/285,477  
PRIOR FILING DATE: 2001-04-20  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 14  
LENGTH: 394  
TYPE: PRN  
ORGANISM: P. aeruginosa  
US-10-125-692-14

Query Match 4.4%; Score 102; DB 9; Length 394;

Best Local Similarity 21.3%; Pred. No. 1.4;  
Matches 71; Conservative 39; Mismatches 92; Indels 132; Gaps 13;

QY 11 IYLAITSAKHGRSDVNSLITKVTALKKASGN-EALALEGLKGTQTPDRVGOAT 69  
DB 72 ISLAQT-ABGALQOQSTNIIQRRMDLSLOSANGSNSDSERTALNGEAKQKELDRISNT 130  
QY 70 -----KIL-GSVGSLGKLSNGDATKII-----SG 93  
DB 131 TFGKRLDGSFQVASFQVQSA-ANETISVIGIDENSAESLNGTYKADGGAVTATAG 189  
QY 94 CEDIYAGIATTEGPGVM----- 111  
DB 190 TVDIALGT--TGSANVNVVDKMGNETAQAATAANVDANGIGARSDGDTISYVK 247  
QY 112 ---GIGAVASFVSSILSLFTGSSA-----KNSVAVID---RALSXH 147  
DB 248 AKDGSAGITSAVSGVSIADTGTGVTAGVAPATAFAKTNIDYAKIDISTAKLSRR 307  
QY 148 ---RDEAIQRHAGAKRDPFAESSAFLOVKKOOSN-----LTDSDLSITANV 193  
DB 308 AGDRTTALIKQIDASVPTSVAAVONRFDNTNLTNLTGENVSAARGRIEDTFAETANL- 365  
QY 194 YKFSNFIGLESRIQGAATTSLSDAKRAVDLIL 227  
DB 366 -----TKNQVLQAGTALIAQANQLPOSVL 390

RESULT 12  
US-09-820-843A-50  
Sequence 50, Application US/09820843A  
Publication No. US2003003963A1  
GENERAL INFORMATION:  
APPLICANT: Council of Scientific and Industrial Research  
TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PF  
FILE REFERENCE: 063915  
CURRENT APPLICATION NUMBER: US/09/820,843A  
PRIOR FILING DATE: 2001-03-30  
NUMBER OF SEQ ID NOS: 118  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 50  
LENGTH: 444  
TYPE: PRN  
ORGANISM: C. pneumoniae CWL029  
FEATURE:  
NAME/KEY: misc-feature  
OTHER INFORMATION: CT579 hypothetical protein  
NAME/KEY: misc-feature  
OTHER INFORMATION: g14377120  
US-09-820-843A-50

Query Match 4.3%; Score 100; DB 9; Length 444;  
Best Local Similarity 24.7%; Pred. No. 2.5;  
Matches 54; Conservative 34; Mismatches 89; Indels 42; Gaps 8;

QY 28 SILKVTALKKASGN-EALALEGLKGTQTPDRVGOATKILGSVSGALGKLSNGDA 87  
DB 143 ALLGLVNTLANNAG-----ESWKASQOSONAIRQOVESAPRIGETIKRQAHQA 193  
QY 88 T-----KIISGLDIYAGIATTEGPGVMGT-----GAVASFVSSILSLFTGSSAKN 134  
DB 194 SATPAQKOSLISGIVNG-----FTVSVGAGIFSAKAKATSAKRSAPKEGASAG 248  
QY 135 SVAAYIDRALSKHDEAIQRHAGAK-RDPAESSAFLOVKKOOSNLTDSDLSITANV 193  
DB 249 GAAS---KALTSASSVQQTMASTAKATTAASSAGSAAATKAANLTD-DMAAASAKMAS 304  
QY 194 YKFSNFIGOL-----ESRISGGAATTSLSDAKRA 222  
DB 305 DGASKAGGGLFGEVLTNKPNNSEKYSRGMNVYKQGANVA 343

PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 5883  
LENGTH: 837  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5883

Query Match 4.5%; Score 104.5; DB 10; Length 837;  
Best Local Similarity 19.1%; Pred. No. 2.5;  
Matches 75; Conservative 71; Mismatches 147; Indels 99; Gaps 17;

QY 16 TSAHGRSDVNSLTETKETAKEASGN-----EALALEGLKGEIQ----- 59  
DB 301 TDADAKRNATYNAVTOAEQILNKAGPNTSKDGVETALENQRANKNELNGNQNANAKT 360  
QY 60 TKPRVGAATKILGSVSGALGKNSGDATKIIISGLDIYAGIATTFGGPVGMIGAVASF 119  
DB 361 TAKNALNLTSINNAQKALKSQIEGAT-----VAGVNO-----VSTTASE 402  
QY 120 VSSILS-LEFTSSAKNSAAVVID-----RALSKHDEAIQ--RHAAGAKRDFAESAFIQ 171  
DB 403 LNTAMSNLQNGINDEAAKALNGTONLEKAKQHANTAIIDGLSHLTNAQKE-----ALKQ 457  
QY 172 VMKQOOSNLTDSDLSIIANPVYKFSNFIQLESRIISOGAAT-----TSLSDA-----KRAVD 224  
DB 458 LVQOSTYAEAGNQRKAN-----NDAMMDKLRQSIADNATTKQONNTDASQNKKDAYN 513  
QY 225 FILYQCQLVNRETL-LVDLAILYR-----KQNAEHVASAVENA 262  
DB 514 NAVTAAQGIIDQTSPTLDPTVINOAGOVSTTKNALNGENLEAKQAQASQSLSDNL 573  
QY 263 NRVKELAADPLDPLHLK-----IPEQA-----LIGAYHPIASSETSKAILNTKYICV 312  
DB 574 NNAQQTVDINGAHYVDEANQIKQONQONNTAMGNLKOALADDAKATVNF----- 628  
QY 313 PDVPRIGNRKRYKFTNSYNTSICSEAYMGN 344  
DB 629 -DADQA---KQAYTAVTNAENITSKANGN 656

RESULT 9  
US-09-815-242-13080  
Sequence 13080, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlssen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 13080  
LENGTH: 875  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-13080

Query Match 4.5%; Score 104.5; DB 10; Length 875;  
Best Local Similarity 19.1%; Pred. No. 2.7;  
Matches 75; Conservative 71; Mismatches 147; Indels 99; Gaps 17;

QY 16 TSAHGRSDVNSLTETKETAKEASGN-----EALALEGLKGEIQ----- 59  
DB 338 TDADAKRNATYNAVTOAEQILNKAGPNTSKDGVETALENQRANKNELNGNQNANAKT 397  
QY 60 TKPRVGAATKILGSVSGALGKNSGDATKIIISGLDIYAGIATTFGGPVGMIGAVASF 119  
DB 398 TAKNALNLTSINNAQKALKSQIEGAT-----VAGVNO-----VSTTASE 439  
QY 120 VSSILS-LEFTSSAKNSAAVVID-----RALSKHDEAIQ--RHAAGAKRDFAESAFIQ 171  
DB 440 LNTAMSNLQNGINDEAAKALNGTONLEKAKQHANTAIIDGLSHLTNAQKE-----ALKQ 494  
QY 172 VMKQOOSNLTDSDLSIIANPVYKFSNFIQLESRIISOGAAT-----TSLSDA-----KRAVD 224  
DB 495 LVQOSTYAEAGNQRKAN-----NDAMMDKLRQSIADNATTKQONNTDASQNKKDAYN 550  
QY 225 FILYQCQLVNRETL-LVDLAILYR-----KQNAEHVASAVENA 262  
DB 551 NAVTAAQGIIDQTSPTLDPTVINOAGOVSTTKNALNGENLEAKQAQASQSLSDNL 610  
QY 263 NRVKELAADPLDPLHLK-----IPEQA-----LIGAYHPIASSETSKAILNTKYICV 312  
DB 611 NNAQQTVDINGAHYVDEANQIKQONQONNTAMGNLKOALADDAKATVNF----- 665  
QY 313 PDVPRIGNRKRYKFTNSYNTSICSEAYMGN 344  
DB 666 -DADQA---KQAYTAVTNAENITSKANGN 693

RESULT 10  
US-09-815-242-5639  
Sequence 5639, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlssen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23

SOFTWARE: Patentin version 3.0

PAKISTAN  
 PAKISTAN APPLICATION NUMBER: 60/242,5/8  
 PRIOR FILING DATE: 2000-10-23



Db 1863 NEAKNTAKOQVATMSHLTDQAKANLTS---QIESG---TTVAGYOGI-----1903  
Qy 126 LETGSSAKNSVAVIDRALSKHDEAIORHAAGAKRDEFAESSAFIQVMKQOSWLDSDS 185  
Db 1904 -----QANAGTLDQANQLOQSASIDATKSSSEDYODANADLONAYADA-VTNAEGI 1954  
Qy 186 IIAANVPYKFSNFIGLESRSISOGAATSLSDAKRAVDIFILYCOLVWRELLVDLAI 245  
Db 1955 ISATNPEKN-PDTINOKASOVN--SAKSALNDEK-----LAAAKQAKSDIGR 2001  
Qy 246 LYRKGNAEHVASAVENANVKNELADTDLFLKLIPEOALIGAVYHPISASETSKAILN 305  
Db 2002 LLDNNNAOQTANAAEVDQAPN--LAAVTA-AKKKATSLMTAMGNLHALAKXONTFRSVA 2058  
Qy 306 YF 307  
Db 2059 YF 2060

RESULT 4  
US-09-841-132-177  
; Sequence 177, Application US/09841132  
; Patent No. US20020061848A1  
; GENERAL INFORMATION:  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Skeiky, Yaser A.W.  
; TITLE OF INVENTION: PROBST, PETER  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; FILE REFERENCE: 210121.469C8  
; CURRENT APPLICATION NUMBER: US/09/841,132  
; CURRENT FILING DATE: 2001-04-23  
; NUMBER OF SEQ ID NOS: 599  
; SOFTWARE: FASTSEQ for Windows Version 3.0/4.0  
; SEQ ID NO 177  
; LENGTH: 964  
; TYPE: PRF  
; ORGANISM: CHLAMYDIA  
US-09-841-132-177

Query Match 4.8%; Score 112.5; DB 10; Length 964;  
Best Local Similarity 21.3%; Pred. No. 0.64;  
Matches 110; Conservative 61; Mismatches 184; Indels 161; Gaps 28;

Qy 21 GKRSDV---NSLITKVEALKEASGSENEALBEGLEIOTKFDYRGQATKILGSGV 76  
Db 254 GKRGNIVFYNNRCFKNVEFASSEAS-----DGAIKYTRLDVYGNRGRIFFS-- 301  
Qy 77 SALGRKNSGDATKIIISGCLDIYAGIATTFGGPYGMGIGAVASPVSSILSFTGSSAKNSV 136  
Db 302 -----DNITKNYGGAI--YAPVYLVLDNGPPTFFINNINANKG 336  
Qy 137 AAVIDRALSKHDEAIORHA-----AGAKR---DFAESSAFIQV 172  
Db 337 GAIYIDGTNSKRISA-DRAHAIIFENIYNTVNTANGTSANPFRNATIVASSGEILL 395  
Qy 173 MKQOS-NLTDSD-LSIIANVPYKFSNFIGLESRSISOGAATSLSDAKRAVD----- 224  
Db 396 GAGSSONLIFYDPIEVSNAQVSF-SFNKEADQTSYVFSGATVNSADFHQRNLQOTKTPAP 454  
Qy 225 -----FILL--YCOLVWRELL--LVLDL--AIL--YRKGNAEHVASAVENANVKNKE 268  
Db 455 LTLNSGFLCIEDHQAOLTVNRFTOTGGVSLGNGAVLSCTYKNGTGD-----SASNASITLKH 510  
Qy 269 LAADTDLFLKLIPEOALIGAVYHPISASETSKAILNT---KYRGPDPVPRPI---G 320  
Db 511 IGIN-----LSIILKSGAEI-----PLWVEPTNNSNNYTAADTAATFSLSDVKSLDDYX 561  
Qy 321 NRRYKFTNSYMTYSICSEAYMGNTYFRGCSNVNRNIRVSKMSDGFYTMENSDRRKLYI 380  
Db 562 NSPYESTDL--THALSSQ-----PMLTISEASDNQLOSENIDFSGLVN 602  
Qy 381 TKHD-OG---WGWG-TLDEDDPGDOGHMR-----FIPL-RHCKYV 414

Db 603 PHYGMQGLTWGAKTQDEPPASSAFTTPOKANREHRTLLTWLPAGVVPSPKHSRPLI 662  
Qy 415 SSKRPMNPMYESSASGYIRSMENPPGQGHST 450  
Db 663 ANTLMGNMLLATESILKN-----SAILTPSGHPFWGIT 694

RESULT 5  
US-09-841-132-191  
; Sequence 191, Application US/09841132  
; Patent No. US20020061848A1  
; GENERAL INFORMATION:  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Skeiky, Yaser A.W.  
; TITLE OF INVENTION: PROBST, PETER  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; FILE REFERENCE: 210121.469C8  
; CURRENT APPLICATION NUMBER: US/09/841,132  
; CURRENT FILING DATE: 2001-04-23  
; NUMBER OF SEQ ID NOS: 599  
; SOFTWARE: FASTSEQ for Windows Version 3.0/4.0  
; SEQ ID NO 191  
; LENGTH: 977  
; TYPE: PRF  
; ORGANISM: CHLAMYDIA  
US-09-841-132-191

Query Match 4.8%; Score 112.5; DB 10; Length 977;  
Best Local Similarity 21.3%; Pred. No. 0.65;  
Matches 110; Conservative 61; Mismatches 184; Indels 161; Gaps 28;

Qy 21 GKRSDV---NSLITKVEALKEASGSENEALBEGLEIOTKFDYRGQATKILGSGV 76  
Db 267 GKRGNIVFYNNRCFKNVEFASSEAS-----DGAIKYTRLDVYGNRGRIFFS-- 314  
Qy 77 SALGRKNSGDATKIIISGCLDIYAGIATTFGGPYGMGIGAVASPVSSILSFTGSSAKNSV 136  
Db 315 -----DNITKNYGGAI--YAPVYLVLDNGPPTFFINNINANKG 349  
Qy 137 AAVIDRALSKHDEAIORHA-----AGAKR---DFAESSAFIQV 172  
Db 350 GAIYIDGTNSKRISA-DRAHAIIFENIYNTVNTANGTSANPFRNATIVASSGEILL 408  
Qy 173 MKQOS-NLTDSD-LSIIANVPYKFSNFIGLESRSISOGAATSLSDAKRAVD----- 224  
Db 409 GAGSSONLIFYDPIEVSNAQVSF-SFNKEADQTSYVFSGATVNSADFHQRNLQOTKTPAP 467  
Qy 225 -----FILL--YCOLVWRELL--LVLDL--AIL--YRKGNAEHVASAVENANVKNKE 268  
Db 468 LTLNSGFLCIEDHQAOLTVNRFTOTGGVSLGNGAVLSCTYKNGTGD-----SASNASITLKH 523  
Qy 269 LAADTDLFLKLIPEOALIGAVYHPISASETSKAILNT---KYRGPDPVPRPI---G 320  
Db 524 IGIN-----LSIILKSGAEI-----PLWVEPTNNSNNYTAADTAATFSLSDVKSLDDYX 574  
Qy 321 NRRYKFTNSYMTYSICSEAYMGNTYFRGCSNVNRNIRVSKMSDGFYTMENSDRRKLYI 380  
Db 575 NSPYESTDL--THALSSQ-----PMLTISEASDNQLOSENIDFSGLVN 615  
Qy 381 TKHD-OG---WGWG-TLDEDDPGDOGHMR-----FIPL-RHCKYV 414  
Db 616 PHYGMQGLTWGAKTQDEPPASSAFTTPOKANREHRTLLTWLPAGVVPSPKHSRPLI 675  
Qy 415 SSKRPMNPMYESSASGYIRSMENPPGQGHST 450  
Db 676 ANTLMGNMLLATESILKN-----SAILTPSGHPFWGIT 707

RESULT 6  
US-09-841-696-5  
; Sequence 5, Application US/09884696

Matches 64; Conservative 58; Mismatches 121; Indels 59; Gaps 12;

QY 16 TSAHKGRSDVNSLTITKVTALKKASGSNE--AALE-----ALBGLKEIQTKPDV 65  
Db 1029 TDADREKOTAYNDVTAKTLLDKTAGSNDKKAVEQALQVNTAKTALND-----ERL 1093

QY 66 GOATKILGSVSGALGKNSGDATKTIISGCLDIYAGIATFTGGPYMGIGAVASVSSILS 125  
Db 1084 NEAKNTAKQOOLATWSHILTNOKANLTS--QIERG--TTVAGVQGI----- 1124

QY 126 LFTGSSAKNSVAIVIDRALSKHRDEAIQRHAGAKKRDFASSAFIYVMOOSNLTDSDL 185  
Db 1125 -----QANACTLDQAMQRLQSTASKKATSSSEDIYQANADLQNAVHA--VTNAEGI 1175

QY 186 ITAANPVYKFSNFICQLESRIQGAATTSLSQAKRAVDFILLYCOLVNRRETLVDLAI 245  
Db 1176 ISATNNPEMN-PDTINOKASQVN--SAKSALNGDEK-----LAAKQSAKTIDGH 1222

QY 246 LYKGAHEHVASAVENANRNKRELADTDLFLKLIPEQALIGAVYHPISASETSKAILN 305  
Db 1223 LTPDLNNAQOTSAVAEVDQAPN--LAAVTA--AKNKATSLNTAMGNLKHALLAEKDNTRKRSVN 1279

QY 306 YT 307  
Db 1280 YT 1281

RESULT 2  
US-09-815-242-12610  
; Sequence 12610, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA 011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12610  
; LENGTH: 5795  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12610

Query Match 5.1%; Score 118.5; DB 10; Length 5795;  
Best Local Similarity 20.9%; Pred. No. 2.4; Mismatches 121; Indels 59; Gaps 12;  
Matches 63; Conservative 59; Mismatches 121; Indels 59; Gaps 12;

QY 16 TSAHKGRSDVNSLTITKVTALKKASGSNE--AALE-----ALBGLKEIQTKPDV 65  
Db 1029 TDADREKOTAYNDVTAKTLLDKTAGSNDKKAVEQALQVNTAKTALND-----ERL 1093

Db 4486 TDADREKOTAYNDVTAKTLLDKTAGSNDKKAVEQALQVNTAKTALNGDA-----RL 4540

QY 66 GOATKILGSVSGALGKNSGDATKTIISGCLDIYAGIATFTGGPYMGIGAVASVSSILS 125  
Db 4541 NEAKNTAKQOOLATWSHILTNOKANLTS--QIERG--TTVAGVQGI----- 4581

QY 126 LFTGSSAKNSVAIVIDRALSKHRDEAIQRHAGAKKRDFASSAFIYVMOOSNLTDSDL 185  
Db 4582 -----QANACTLDQAMQRLQSTASKKATSSSEDIYQANADLQNAVHA--VSDAEGI 4632

QY 186 ITAANPVYKFSNFICQLESRIQGAATTSLSQAKRAVDFILLYCOLVNRRETLVDLAI 245  
Db 4633 ISATNNPEMN-PDTINOKASQVN--SAKSALNGDEK-----LAAKQSAKTIDGH 4679

QY 246 LYKGAHEHVASAVENANRNKRELADTDLFLKLIPEQALIGAVYHPISASETSKAILN 305  
Db 4680 LSPDLNNAQOTSAVAEVDQAPN--LAAVT--SAKNKATSLNTAMGNLKHALLAEKDNTRKRSVN 4736

QY 306 YT 307  
Db 4737 YT 4738

RESULT 3  
US-09-815-242-12996  
; Sequence 12996, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA 011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12996  
; LENGTH: 6281  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12996

Query Match 5.1%; Score 118.5; DB 10; Length 6281;  
Best Local Similarity 21.2%; Pred. No. 2.7; Mismatches 121; Indels 59; Gaps 12;  
Matches 64; Conservative 58; Mismatches 121; Indels 59; Gaps 12;

QY 16 TSAHKGRSDVNSLTITKVTALKKASGSNE--AALE-----ALBGLKEIQTKPDV 65  
Db 1808 TDADREKOTAYNDVTAKTLLDKTAGSNDKKAVEQALQVNTAKTALND-----ERL 1862

QY 66 GOATKILGSVSGALGKNSGDATKTIISGCLDIYAGIATFTGGPYMGIGAVASVSSILS 125  
Db 1029 TDADREKOTAYNDVTAKTLLDKTAGSNDKKAVEQALQVNTAKTALND-----ERL 1093

GemCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2003, 11:17:28 ; Search time 57 Seconds  
(without alignments)  
761.454 Million cell updates/sec

Title: US-09-647-522-5

Perfect score: 2332  
Sequence: 1 MILKHPWLFVLTATSAKH.....SGYISWENNEPQCHMSIT 450

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 362588 seqs, 96450795 residues

Total number of hits satisfying chosen parameters: 362588

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubppaa/PCCT\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubppaa/PCCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubppaa/PCCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118.5	5.1	2434	10 US-09-815-242-5835	Sequence 5835, Ap
2	118.5	5.1	5795	10 US-09-815-242-12610	Sequence 12610, A
3	118.5	5.1	6281	10 US-09-815-242-12996	Sequence 12996, A
4	112.5	4.8	964	10 US-09-841-132-177	Sequence 177, App
5	112.5	4.8	977	10 US-09-841-132-191	Sequence 191, App
6	108.5	4.7	1023	9 US-09-884-696-5	Sequence 5, Appl
7	105.5	4.5	2504	10 US-09-817-514A-8	Sequence 8, Appl
8	104.5	4.5	837	10 US-09-815-242-5883	Sequence 5883, Ap
9	104.5	4.5	875	10 US-09-815-242-13080	Sequence 13080, A
10	104.5	4.5	2086	10 US-09-815-242-5639	Sequence 5639, Ap
11	102	4.4	394	9 US-10-125-692-14	Sequence 14, Appl
12	100	4.3	3158	9 US-09-820-843A-50	Sequence 50, Appl
13	100	4.3	3158	9 US-09-815-242-12611	Sequence 12611, A
14	98.5	4.2	379	9 US-10-125-692-13	Sequence 13, Appl
15	97	4.2	2122	9 US-09-813-214A-9	Sequence 9, Appl
16	95	4.1	434	9 US-09-892-877-459	Sequence 459, App
17	94	4.0	722	10 US-09-815-242-12888	Sequence 12888, A
18	94	4.0	991	10 US-09-815-242-5803	Sequence 5803, Ap
19	94	4.0	1215	10 US-09-815-242-5908	Sequence 5908, Ap

20	94	4.0	1269	10 US-09-815-242-13113	Sequence 13113, A
21	94	4.0	2437	10 US-09-815-242-5834	Sequence 5834, Ap
22	93.5	4.0	429	10 US-09-815-242-13506	Sequence 13506, A
23	93	4.0	1295	10 US-09-726-949A-1	Sequence 1, Appl
24	92.5	4.0	432	10 US-09-919-172-9	Sequence 9, Appl
25	92.5	4.0	459	9 US-10-102-806-469	Sequence 469, App
26	92	4.0	3241	10 US-09-841-786-1	Sequence 1, Appl
27	91	3.9	1043	10 US-09-815-242-5194	Sequence 5194, Ap
28	91	3.9	1289	9 US-09-712-363-259	Sequence 259, App
29	91	3.9	2025	10 US-09-815-242-5703	Sequence 5703, Ap
30	90.5	3.9	550	9 US-10-125-692-22	Sequence 22, Appl
31	90.5	3.9	554	9 US-10-125-692-21	Sequence 21, Appl
32	90	3.9	375	9 US-09-738-628-6333	Sequence 6333, Ap
33	90	3.9	420	9 US-09-770-509-10	Sequence 10, Appl
34	89.5	3.9	483	8 US-08-834-656A-20	Sequence 20, Appl
35	89.5	3.9	628	10 US-09-881-736-4	Sequence 4, Appl
36	89.5	3.9	660	12 US-10-007-693-139	Sequence 139, App
37	89	3.8	89	10 US-09-925-302-445	Sequence 445, App
38	89	3.8	803	10 US-09-738-363-12	Sequence 12, Appl
39	88.5	3.8	4613	9 US-09-860-846-31	Sequence 31, Appl
40	88.5	3.8	4613	9 US-09-988-384B-31	Sequence 31, Appl
41	88.5	3.8	4613	9 US-09-836-821-31	Sequence 31, Appl
42	88.5	3.8	4613	10 US-09-861-289-31	Sequence 31, Appl
43	88.5	3.8	11877	9 US-09-860-846-6	Sequence 6, Appl
44	88.5	3.8	11877	9 US-09-836-821-6	Sequence 6, Appl
45	88.5	3.8	11877	10 US-09-861-289-6	Sequence 6, Appl

#### ALIGNMENTS

RESULT 1  
US-09-815-242-5835  
Sequence 5835, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsson, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 5835  
LENGTH: 2434  
TYPE: prt  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5835  
Query Match 5.1%; Score 118.5; DB 10; Length 2434;  
Best Local Similarity 21.2%; Pred. No. 0.72;



OY 184 LSI-----IAANVP-----YKFSNFIGLESRI-----SOGAA 212  
DB 890 MSAFEGEITPLVPSDEKRLYMGTRNDFTTYAGCGGTRPKQOYIADHTQVCSANVM 949  
OY 213 TTSISDAKRAVDPI-----LLYCOLVMRETLVLDAILIRKGNABHV 255  
DB 950 YTALSRAITDRHFVNTSANSASFWEKLDSTPYLTKFLSVNEQALRE---YEPAAEAPI 1005  
OY 256 AS-----AVENANRVKELAADTL-----FLHKIIP 282  
DB 1006 REPEPOTMCVENESEVLEEKKEELLEKFREIHSESHGNCVOTEDTTQLSHQOAK 1065  
OY 283 EOALIGAVYH---PISASETS-----KALINVTYFGVDPVRPDRIGNRRYFT 327  
DB 1066 DETILMATIDARKLISNOETFRFELSKDIDGLVPLFYQAKMLP-----KERIPFS 1118  
OY 328 NSYWNYSICSEAYMNGVMFRGCSNVRNPNIIVSKMSDGFYTMENSDBRKLITRKHDGW 387  
DB 1119 QEVWEA---CAHEVQSKYLSKSCNLIINGTVROS-----PDFENKIMVEFLKSQ-- 1164  
OY 388 GWCGLDEDPDQGHMRFP-LRHGKYW-----SKRRPNMFMWESSASGYIR 435  
DB 1165 -WYTKVYKRLG-----LPKIKPGOTIAAFYQOTVLMFGTMARYMFRORAFOPREYFI- 1215  
OY 436 SWENNPGRPOGHWSI 449  
DB 1216 NCETTPEDMSKML 1229

## RESULT 12

US-09-134-001C-5282  
; Sequence 5282, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5282  
; LENGTH: 331  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5282

Query Match 4.18; Score 94.5; DB 4; Length 331;

Best Local Similarity 23.0%; Pred. No. 0.28;

Matches 80; Conservative 50; Mismatches 111; Indels 107; Gaps 21;

OY 10 FIVLA-ITSAHGRSDVNSLITKVEALKEASGS-----NE-----AALEALE 52  
DB 13 FIMKYIDEQTOALDNEVILEVERALQAFSEKKTITPRIVLPFEQNRNRYLVMPALS 72  
OY 53-----GLKEIQRKPD--RVGAQTKILGSV-----GSALGKLSNGDATKIISGCDI 97  
DB 73 DELNIVGLK-IVSFAPENSKKAT-ITGSVILSDYETGELSLIDGFLTKVRGA--- 127  
OY 98 VAGIATTFGGPVGMIGAVASFVSSILSF-TGSSAKNSVAAYVD-RALSXHRDAIORH 155  
DB 128 ISGATRY-----LAKENAKTLVIGAGVQAEGLIEAIIAIVRDIER-----IH 170  
OY 156 AAGAKRDEASSAFIYVKKOOSN-----LTPDSLSITA--ANVPYKFS----- 197  
DB 171 I--ASRFEKAEKFAQNRNFRNFKVSVFRSADAEIDSDADIVYKATNANOYTHSLHPG 228  
OY 198 ---NFIOLESRIISOGAATTSLSDAKRAVDLILYCOLVMRETLVLDAILYRKG--- 250

DB 229 VHLNAGSFKPDQOIPSETWLVANK-----IVSEMEALEET--GDLKIPQAGILTK 281  
OY 251 -----NAHVASAVENANRVKELAADL-----FLHKIIP 281  
DB 282 NMLHSELGDIISGERVRETEEVYKRSVGLAIVDIIIVAOYFKKLI 329

## RESULT 13

5183745-2  
; Patent No. 5183745  
; APPLICANT: DANCHIN, ANTOINE; GLASER, PHILIPPE; KRIN, EVELYN;  
; BARZU, OCTAVIEN; LADANT, DANIEL; OULMAN, AGNES  
; TITLE OF INVENTION: ADENYL CYCLASE DERIVATIVES AND THEIR  
; BIOLOGICAL USES  
; NUMBER OF SEQUENCES: 13  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/426,541  
; FILING DATE: 25-OCT-1989  
; SEQ ID NO: 2  
; LENGTH: 1489  
5183745-2

Query Match 4.18; Score 94.5; DB 6; Length 1489;

Best Local Similarity 25.0%; Pred. No. 3.6;

Matches 65; Conservative 49; Mismatches 115; Indels 31; Gaps 12;

OY 11 IVLAITSKHKRSDVNSLITKVEALKEASGSMEALEGLKEIQ-----TKPRV 65  
DB 587 IALALAAAR-----GVTSGLVAGASAGAAALALALSPME-TYGLVQOSHVADOLDKL 640  
OY 66 GOATRIIGSVGSLKLSNGDATKIISGDIIVAGIATTFGPGVMGIGA--VASFVSSI 123  
DB 641 ADESSANGYEDDALLAOLYRKPT-AABCAVAGVAVLSTVGAANSIAAASVYGAFAVY 699  
OY 124 LSEFTGSARKNSVAVIDRALSKEH--RDEAIOHAAKARDFAESSAFIYVKKOOSMLTD 181  
DB 700 TSLTIG--ALNGILRGVQOPIETKLANDYARKIDELGGQAYFEKN--LQARHQLANSD 755  
OY 182 SDSLITANVYVYKFSNFIGLESRIISG-----AATTSLSAKRAVD-FILLYCO-LVYM 235  
DB 756 GLRRMLADLQAGWNASVTVQOTTEISKSALELAITGNADLNKSDVDFRFGGERVA 815  
OY 236 RETLLVDA-----ILYRK 250  
DB 816 GQPVLDVYAAAGIDIASRKG 835

## RESULT 14

US-08-669-785-2  
; Sequence 2, Application US/08669785  
; Patent No. 6309648  
; GENERAL INFORMATION:  
; APPLICANT: Betsou, Fotini  
; APPLICANT: Sebo, Peter  
; APPLICANT: Guiso, Nicole  
; TITLE OF INVENTION: Protective Epitopes Of Adenyl  
; TITLE OF INVENTION: Cyclase-Haemolysin(AC-Hly), Their Application To  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flunegan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/669,785

APPLICANT: Turner, Nilgun E.  
TITLE OF INVENTION: Plants Resistant to Infection by PVX  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. B44F  
STREET: 700 Chesterfield Village Parkway  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63198

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/803,973  
FILING DATE: 21-FEB-1997  
CLASSIFICATION: 800

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/804,862  
FILING DATE: 04-DEC-1991  
APPLICATION NUMBER: US 07/771,912  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Hoerner Jr., Dennis R.  
REGISTRATION NUMBER: 30,914  
REFERENCE/DOCKET NUMBER: 38-21(10541)A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (314)537-6099  
TELEFAX: (314)537-6047  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1456 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-803-973-2

Query Match 4.1%; Score 95; DB 1; Length 1456;  
Best Local Similarity 19.0%; Pred. No. 3.1;  
Matches 105; Conservative 67; Mismatches 188; Indels 194; Gaps 27;

QY 52 EGLKEIQTKPDVGOATKILGSVS-----ALGKINSQ-DATKIISG---CLDIYA 99  
Db 714 EYLTKIESLTERKVATCVIHGAGSGSKSHAIQKALREIGKSDITVVLPTNELRLDWSK 773  
QY 100 GIATP-----FGG-----PYGMGIGAVASFYSSI-LSLFTGSSAK 133  
Db 774 KVPNTEPYMFKTEKALIGTGSIVIPDDYSKLPFGY-IEALICFYSKIKLIVLTGDSRQ 832  
QY 134 NSVAVIDRALSKRDEAIO-----RHAAGA---KRDFAESSAFIQVMKQSNLTDSD 183  
Db 833 SVYHETADASIRHLGPATEYFSKYCRYLYNATHRNKRLDAN--MLGVYSERTGVEIS 889  
QY 184 LSI-----IAANVP-----YKFSNFIQLESRI-----SQGA 212  
Db 890 MSAELBESITPLVPSDEKRLKLYMTGTRNDFTYAGCGGLTRPKVQIVLDHNTQVCSANVM 949  
QY 213 TTSISDAKRAVDFT-----LYCOLVVMRETLVLDLAILYRKNANAHV 255  
Db 950 YTALSRAIDRIHFVNTSANSASFMEKLDSTPYLKTFLSVADQALRE---YEPAEAEPT 1005  
QY 256 AS-----AVENANRVKELAAOTLD-----FLAKLIP 282  
Db 1006 REPPOQHMCEVNESEVLEEKLEKFEIRHSESHGNSCVQTEDTTLQPSHOAK 1065  
QY 283 EOALIGAVYH---PISASETS-----KAILNYTYFGVPPDPRIGRRRYFT 327  
Db 1066 DETLLMTADRLKISNOETFRFELSKKIDGLVFLYQKAMGIP-----KERIPFS 1118  
QY 328 NSYVNTYSICSEAVYMGNYMFGCSNVNRPNIRVSKMSDGFYTMENSDBRKLYITRKDOGW 387

Db 1119 OEYWEA---CAHEYQSKYLSKSCNKLINGTVROS-----PDPENKIMVFLKSQ-- 1164  
QY 388 GWTGLDEDDPDGDMFIP-LRHGKIYV-----SSKRPNNFMYMESSASGYR 435  
Db 1165 -WTKYKEKLG-----LPRIKPGQITAFYQOVTMLGTMARYMRFROAFQPEVEFI 1215  
QY 436 SWENNPQPGQHWSI 449  
Db 1216 NCETTPEDMSAMAL 1229

RESULT 11  
US-08-803-972-2  
Sequence 2, Application US/08803972  
Patent No. 5792937  
GENERAL INFORMATION:  
APPLICANT: Braun, Carl J.  
APPLICANT: Hemenway, Cynthia L.  
APPLICANT: Turner, Nilgun E.  
TITLE OF INVENTION: Plants Resistant to Infection by PVX  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. B44F  
STREET: 700 Chesterfield Village Parkway  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63198

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/803,972  
FILING DATE: 21-FEB-1997  
CLASSIFICATION: 800

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/804,862  
FILING DATE: 04-DEC-1991  
APPLICATION NUMBER: US 07/771,912  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Hoerner Jr., Dennis R.  
REGISTRATION NUMBER: 30,914  
REFERENCE/DOCKET NUMBER: 38-21(10541)A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (314)537-6099  
TELEFAX: (314)537-6047  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1456 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-803-972-2

Query Match 4.1%; Score 95; DB 1; Length 1456;  
Best Local Similarity 19.0%; Pred. No. 3.1;  
Matches 105; Conservative 67; Mismatches 188; Indels 194; Gaps 27;

QY 52 EGLKEIQTKPDVGOATKILGSVS-----ALGKINSQ-DATKIISG---CLDIYA 99  
Db 714 EYLTKIESLTERKVATCVIHGAGSGSKSHAIQKALREIGKSDITVVLPTNELRLDWSK 773  
QY 100 GIATP-----FGG-----PYGMGIGAVASFYSSI-LSLFTGSSAK 133  
Db 774 KVPNTEPYMFKTEKALIGTGSIVIPDDYSKLPFGY-IEALICFYSKIKLIVLTGDSRQ 832  
QY 134 NSVAVIDRALSKRDEAIO-----RHAAGA---KRDFAESSAFIQVMKQSNLTDSD 183  
Db 833 SVYHETADASIRHLGPATEYFSKYCRYLYNATHRNKRLDAN--MLGVYSERTGVEIS 889

TITLE OF INVENTION: MORAXELLA CAPARRHALIS OUTER MEMBRANE  
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/968,685A  
FILING DATE: No. 6214981ember 12, 1997  
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Baldwin, Geraldine F.  
REGISTRATION NUMBER: 31,232  
REFERENCE/DOCKET NUMBER: 7969-060  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2123 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-968-685A-10

Query Match 4.2%; Score 97; DB 4; Length 2123;  
Best Local Similarity 21.9%; Pred. No. 3.6; Mismatches 127; Indels 100; Gaps 16;  
Matches 75; Conservative 40;

QY 42 GSENALEALBGLKGEIOTKPPDR---VQATKILGSVGSALG--RL--NSGDATKI---I 91  
DB 70 GISER-----DQKGGANARQDKSTAIGDIALGSGSAIDGNKIVHNSNNANIGAKA 124  
QY 92 SGCLDIYAGIATTFGGPYGMGIGA-----VASFVSSILFTGSSAKNSVAAYIDRA 143  
DB 125 SGNESIAIGGVLAGSHASIAIGSDLYLKKEVVOQISELPIIRGOKALNDIYOLADTN 184  
QY 144 LSKHDEALIORHAAGAKRDFESSAFIQVMKQSNLTDSLSITANVPVYFESNPIQL 203  
DB 185 LQKTRRTAOGHASTA---VGAMSYAKGHFSNAFGTR 218  
QY 204 ESRIQGAATYSLDAKRAVDIFILLYCOLVYARETLVLDLALILYKRNAAEHV---ASAVE 260  
DB 219 ATAEGTYSIAGVLTATARAASIAV-----GSNAQATGFAATANG 258  
QY 261 NANRVNKEIADDTDLFLKHLPEQALIGAVYHPISASSETSALKALNTYKYEVPDVPPIG 320  
DB 259 GSTQVNLNNGI-ALGFGSQVLQKDMNDVNAA-----NVRAV--APDDNQPTD 301  
QY 321 NRRYK--FTNSYWNVTYSICSEAYMGNMYMFRGCSNVRNPNIRY 360  
DB 302 N-RYKATERNKATDVESI-----GN--SNGNDSIRKRLINV 334

RESULT 9  
US-08-989-925-1  
Sequence 1, Application US/08989925  
Patent No. 5989620  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Shah, Purvi

APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: HUMAN ADIPOPHILIN-LIKE PROTEIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/989,925  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PR-0440 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 434 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRATU13  
CLONE: 1620223  
US-08-989-925-1

Query Match 4.1%; Score 95; DB 2; Length 434;  
Best Local Similarity 24.5%; Pred. No. 0.4; Mismatches 94; Indels 62; Gaps 10;  
Matches 62; Conservative 35;

QY 17 SAKHGR-----SDVNSLTQVETALKESGSENALEALBGLKGEIOTKPPDRVQAT 69  
DB 62 AAEKGVRTLTAAVAGAPILSKLEPQIASSEYAHRGIDKLEENLPILQOPTERYVADT 121  
QY 70 KILGSVGSALGKNSGDATKIISGCLDIYAGIATTFGGPYGMGIGAIVASFSILSLFT- 128  
DB 122 KEL--VSSKRV---SG-AQEWVSSAKDV---ATQLSEAVDATRGAVOSGVDKTSVYTG 171  
QY 129 -----GSSAKNSVAVIDRALSKRDEAIOHHAAGAKRDFESSAFIQVMKQSNLTDS 182  
DB 172 GVQSWGSRILGOMVLSGVDYVLGK-SEEMADNHL-----LIDA 209  
QY 183 DLSTIANVPVYKESNF-----ISQLESRIQGAATYSL-----SDAKRAVDFT 226  
DB 210 ELARIATSLDGDVAVSQOQROEQSYFVRLGSLSLRLQHVHSLGKLRATKQRAQEA 269  
QY 227 LLYCOLVYARETL 239  
DB 270 LQLSQALSLMETV 282

RESULT 10  
US-08-803-973-2  
Sequence 2, Application US/08803973  
Patent No. 5773701  
GENERAL INFORMATION:  
APPLICANT: Braun, Carl J.  
APPLICANT: Hemenway, Cynthia L.

APPLICANT: Bloecker, Helmut  
APPLICANT: Brandt, Petra  
APPLICANT: Cino, Paul M  
APPLICANT: Dougherty, Brian A  
APPLICANT: Goldberg, Steven L  
APPLICANT: Hoffe, Gerhard  
APPLICANT: Mueller, Joachim  
APPLICANT: Reichenbach, Hans  
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or  
TITLE OF INVENTION: heteropolypeptide compounds  
FILE REFERENCE: PCT/US 99/23535  
CURRENT APPLICATION NUMBER: US/09/413,814  
CURRENT FILING DATE: 1999-10-07  
EARLIER APPLICATION NUMBER: DE 198 46 493.2  
EARLIER FILING DATE: 1998-10-09  
NUMBER OF SEQ ID NOS: 107  
SOFTWARE: Patentn Ver. 2.1  
SEQ ID NO 80  
LENGTH: 3079  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-413-814-80

Query Match 4.3%; Score 100.5; DB 4; Length 3079;  
Best Local Similarity 22.5%; Pred. No. 2.9; Mismatches 126; Indels 113; Gaps 18;

Matches 83; Conservative 47; Mismatches 126; Indels 113; Gaps 18;

QY 38 KEASGSNEALLEGKGIQTRPDR-----VCOATKILGSGV 76  
DB 55 REAGGAEYWRKALDGTATIDLPDRARADAGARGRAHATLPKLTALRLAERG 114  
QY 77 SALGKLNSGDATKII--SCCLDIYAGI-----ATTFGG----PYGNGIGAVAS 118  
DB 115 TTFEVLKSLATVLRHARSQSDLVVGVPSAGRHDESSARAFYFVOMLFRVALRGAAS 174  
QY 119 FVSSILSLFTGSSAKNSVAVIDRALS--KRDDEATORHAGAKRDEASSAFIYWKQ 176  
DB 175 FDA-----LVARVRDAFLDALAHGDSALRHLLAROGEGORALDVAFAFO 221  
QY 177 SN--LTPDLSITIAAVPYKFSNFTGOLESRTSG--AATTSLSDAKRAVDF--ILLYCO 231  
DB 222 STPPSLDAKLSALA-----IGVGDVRIAGOELELTITLADQAAAEFDLALFAAE 270  
QY 232 L---VVMR---ETLVLDLAILYRKNAER-----VASAVENAVNNKELADTIDFLHKL 281  
DB 271 LDAGIALFREYDQGLDPATIERM--ARHFVULLESVAHEPGPRLSELNM--LSDAERALL 327  
QY 282 -----PEQALIGAVYHPISASETSKAILNTYKYEFGVPDVPPIGNRRYK 325  
DB 328 LDDWSGAAARQASAPAPACVHALFEAHARQPDATALEF-----CHORFT 374  
QY 326 FTN-SYWN 333  
DB 375 YAOJSTWST 383

RESULT 7  
US-09-255-829-18

Sequence 18, Application US/09255829  
Patent No. 6461617  
GENERAL INFORMATION:  
APPLICANT: Shone, Clifford Charles  
APPLICANT: Quinn, Conrad Padraig  
APPLICANT: Foster, Keith Alan  
TITLE OF INVENTION: Recombinant Toxin Fragments  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.  
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/255,829  
FILING DATE: 23-FEB-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB97/02273  
FILING DATE: 22-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/782,893  
FILING DATE: 27-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: ESMOND, ROBERT W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 1581.0130002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1013 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-255-829-18

Query Match 4.2%; Score 97.5; DB 4; Length 1013;  
Best Local Similarity 20.9%; Pred. No. 0.91; Mismatches 154; Indels 163; Gaps 25;

QY 9 LFIIVLAITSKCK-----RSDVNSLL-----TKVET-----ALKEASGSNEAA--LEALRG 53  
DB 557 MEHYLAOEFEHGKSRIALTNSVNEALLNPSRYFFSSDYKVKVKAATEAMFGLGWQ 616  
QY 54 L-----KGEIOTKPDVGOATKILGSGALGKNSGD-----ATKIISGC---L 95  
DB 617 LVYDFDESEVST--TDKADITITIIPIYIGPA--LNIQNMUYKDDFVAGALJFSGAVILL 672  
QY 96 DIVAGIATTEFGPGVGIGAVASFWSSILSFTGSSAKNSVAAV--IDRALSKHRD----- 149  
DB 673 EFIPETAI-----PV-LGTFALVSYIA-----NKVLTVQITIDALSKRNEKWE 715  
QY 150 -----EAIORHAGAKRDE--AESSAFIYWKQOOSNLTDSD 183  
DB 716 YKTYVTNMLAKVNTQIDLIRKKMEALENCAEATKAITIYQNYQTEEEKNNINFNID 775  
QY 184 L-----SIIAANVPYKFSNFI--G--QLESRIISGCAATTSLSDAKRAVDFTILYCOLV 234  
DB 776 LSKKNESINKAMINIKNFLNOCSTYILNNSKIPYG-----VKRLDF----- 818  
QY 235 MRETLVDLAILYRKNAEHAASAVEN--ANRYNKLADTIDFLHKLIPQALIGAVYHP 293  
DB 819 --DASLKDALKIYIDNGLTIGQVDRKDKVNNMTLSDIPOLSKYVDNDRLLSTFEY 876  
QY 294 ISASETSKAILNTYKYEFGVPDVPPIGNRRYKFTNSWNT--YSICSEAYMGNYMFGCSN 352  
DB 877 IKSGINSPGAHYAHQHDAVD-----NKFKEQOQNAFEYI----- 911  
QY 353 VRNNIRYSKMSDGYFMENSDRKLTYTKHDCWGKGTIDEDPDGQGHM 402  
DB 912 LHLPLN-----NEQDNNAFIQ-----SLKDDPSQSANL 939

RESULT 8

US-08-968-685A-10  
Sequence 10, Application us/08968685A  
Patent No. 6214981  
GENERAL INFORMATION:  
APPLICANT: TUCKER, KENNETH  
APPLICANT: PLOSTIA, LAURA





TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF CHRONIC INFLAMMATORY

ESUIT 4  
 S-09-620-412C-191  
 Sequence 191, Application US/09620412C  
 Patent No. 6448234  
 GENERAL INFORMATION:  
 APPLICANT: Steven P. Fling  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
 TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 19, 2003, 11:14:52 ; Search time 30 Seconds  
(without alignments)  
441.344 Million cell updates/sec

Title: US-09-647-522-5

Perfect score: 2322  
Sequence: 1 MLTKHLPWLFYLAITSAKH.....SGYISWENNPGQGHWSIT 450

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/PCRUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/backfillseq1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112.5	4.8	964	4	US-09-556-877-177 Sequence 177, App
2	112.5	4.8	964	4	US-09-620-412C-177 Sequence 177, App
3	112.5	4.8	977	4	US-09-556-877-191 Sequence 191, App
4	112.5	4.8	977	4	US-09-620-412C-191 Sequence 191, App
5	110	4.7	3031	1	US-07-689-008-2 Sequence 2, App1
6	100.5	4.3	3079	4	US-09-413-814-80 Sequence 80, App1
7	97.5	4.2	1013	4	US-09-255-829-18 Sequence 18, App1
8	97	4.2	2123	4	US-08-968-685A-10 Sequence 10, App1
9	95	4.1	1434	2	US-08-989-925-1 Sequence 1, App1
10	95	4.1	1456	1	US-08-803-973-2 Sequence 2, App1
11	95	4.1	1456	1	US-08-803-972-2 Sequence 2, App1
12	94.5	4.1	1456	1	US-08-803-972-2 Sequence 2, App1
13	94.5	4.1	1489	6	US-09-134-001C-5282 Patent No. 5183745-2
14	94.5	4.1	1706	4	US-08-669-785-2 Patent No. 5183745
15	94.5	4.1	1794	6	US-08-669-785-2 Patent No. 5183745
16	93.5	4.0	580	2	US-08-591-079-2 Sequence 2, App1
17	93	4.0	1286	1	US-08-460-604A-28 Sequence 28, App1
18	93	4.0	1286	2	US-08-405-96A-28 Sequence 28, App1
19	93	4.0	1286	2	US-08-405-96A-28 Sequence 28, App1
20	92.5	4.0	432	2	US-08-705-660-18 Sequence 18, App1
21	92.5	4.0	432	3	US-08-989-045-18 Sequence 18, App1
22	92.5	4.0	689	2	US-08-684-865-16 Sequence 16, App1
23	92.5	4.0	689	3	US-08-124-491-16 Sequence 16, App1
24	92.5	4.0	907	4	US-09-255-829-16 Sequence 16, App1
25	92	4.0	497	4	US-08-740-223A-14 Sequence 14, App1
26	92	4.0	497	4	US-09-709-188-14 Sequence 14, App1
27	92	4.0	593	2	US-08-591-079-8 Sequence 8, App1

28	92	4.0	593	2	US-08-591-079-10 Sequence 10, App1
29	91.5	3.9	491	4	US-08-740-223A-13 Sequence 13, App1
30	91.5	3.9	491	4	US-09-709-188-13 Sequence 13, App1
31	91.5	3.9	937	1	US-08-253-155A-31 Sequence 31, App1
32	91.5	3.9	585	2	US-08-426-125-2 Sequence 2, App1
33	91	3.9	585	2	US-08-455-355-2 Sequence 2, App1
34	91	3.9	1026	2	US-08-542-003-6 Sequence 6, App1
35	91	3.9	1026	2	US-08-322-760A-6 Sequence 6, App1
36	91	3.9	1026	4	US-09-236-949-6 Sequence 6, App1
37	90.5	3.9	1705	4	US-08-669-785-4 Sequence 4, App1
38	90.5	3.9	10182	4	US-09-134-001C-3159 Sequence 3159, App1
39	90	3.9	619	1	US-07-762-132A-2 Sequence 2, App1
40	90	3.9	619	1	US-08-301-722A-4 Sequence 4, App1
41	89.5	3.9	535	4	US-09-117-860-18 Sequence 18, App1
42	89.5	3.9	15281	2	US-08-471-119A-2 Sequence 2, App1
43	89	3.8	401	2	US-08-591-079-4 Sequence 4, App1
44	89	3.8	803	1	US-08-158-232-10 Sequence 10, App1
45	89	3.8	803	1	US-08-304-626-10 Sequence 10, App1

# ALIGNMENTS

RESULT 1	US-09-556-877-177	Sequence 177, Application US/0955687
Patent No. 6432916	GENERAL INFORMATION:	
APPLICANT: Probst, Peter	APPLICANT: Bhatia, Ajay	
APPLICANT: Skeiky, Yasser	APPLICANT: Flind, Steve	
APPLICANT: Maisonneuve, Jeff	TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND	
FILE REFERENCE: 210121.469C5	DIAGNOSIS OF CHLAMYDIAL INFECTION	
CURRENT APPLICATION NUMBER: US/09/556,877	CURRENT FILING DATE: 2000-04-19	
NUMBER OF SEQ ID NOS: 305	SOFTWARE: FastSeq for Windows Version 3.0/4.0	
SEQ ID NO 177	LENGTH: 964	
TYPE: PRT	ORGANISM: Chlamydia	
US-09-556-877-177	Query Match	4.8%; Score 112.5; DB 4; Length 964;
	Best Local Similarity	21.3%; Pred. No. 0.023;
	Matches 110; Conservative	61; Mismatches 184; Indels 161; Gaps 28;
QY	21 GKRDV-----NSLTKVETALKEAGSGNEALELDELKGEIQKPPRGQATKILGSGV 76	
DB	254 GNRGNIEFYNNRCKPKNETASSSENS-----DCGALKVTRLDVGTGRGIFPS-- 301	
QY	77 SALGKLSGDKATKISGCLDVIAGIATTFGCGVGMIGGAVAFSSILSFTGSSAKNSV 136	
DB	302 -----DNITKNYGCAT---YAPVYTLVDNCPYFINNIANNKG 336	
QY	137 AAVIDRALSKRDEALQIRHA-----AGAKR-----DFAESSAFIOV 172	
DB	337 GAIYIDGTSNKSISA-DRHAIFENETIVTNTANGTSTSNPFRRAIVTAASSGCEILL 395	
QY	173 MKQOS-NUTBSD-LSIITANVPYKFSNFIQLESRSOGAATLSLDARAVD----- 224	
DB	396 GAGSSQNIIFYPDPLEVNAGVSV-SFNKEDQTSVVFSGATVNSADPHORNLQTKTPAP 454	
QY	225 -----FILL--YQOLVMEETL---LYDL---AIL--YKGNAEHVASAVENANRYNKE 268	
DB	455 LITLSNGFLCIEDHQLVNNRTQGTGVVSLGNGAVLSCYKNGTID-----SASNASITLKH 510	
QY	269 LAADTLDFLKLIDPQALIGAVYHPISASSETSKAILNT---KYFGVPDVPRI---G 320	
DB	511 IGLN-----LSSILKSGAEI-----PLMWEPNTNNSNYTADTAATFSISLIDVYKSLIDY 561	

Db 492 ICSDRTGT-----LTLNOMTVV-ET-----YAGSKMDVA---DNPSGLHPKLVV--- 532  
OY 275 DELHKLIFE---QALIGAVHP-----ISASETSKALINTKTYFGV 312  
Db 533 -----LISEGVAQNTTGNIFHPKDGGEVEISGSPTEKAILSWAYKLGW 575

Search completed: May 19, 2003, 11:16:24  
Job time : 87 secs

PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140685.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152353.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161992.  
PR 29-OCT-1999; 99US-0162142.

Query Match 4.78; Score 110; DB 21; Length 1086;  
Best Local Similarity 22.88; Pred. No. 2.4; Mismatches 118; Indels 146; Gaps 20;  
Matches 93; Conservative 51;

QY 9 LFVIAITSAKHKRSD-----VNSLTKVEPAKBSGSENALEBGLK 55  
DB 210 ILITAAVTSIALGKTEGKRGMDGGSIAFAVLIVTVAVSPYRSLO--PQNLDER 267  
QY 56 GEIQFKPDVRGQATKI-----LGSVSGALGKINSQDATKI----- 90  
DB 268 RNIOLEVMRGGRVTKISIDYDVGVPIPLRIGDVPADGVLSHSLAIDESSMTGESKI 327  
QY 91 -----ISGC-----LDIYAGIATFG-----GPGMGIGAVAS 118  
DB 328 VHKDOKSPFLMSGCKVADGVGNMLVGTGVINTFGLMASISEDTGETPLQVRLNGIAT 387  
QY 119 FVSSILSLFTGSSAKNSVAVIDRALSKHRDEAIORHAAGAKRDPFASSAFIQVKKOOSN 178  
DB 388 FI-GIVGL-----SVALVIVLAL-----LVRYFTGTOTDPTNATQIKTTSISD 431  
QY 179 LTDS-----DLSITANP-----YKFSNFIGQ--LESRIS---OGAATT 214  
DB 432 IVDCKVKIFITIAVIVVAVVEGLPLAVTTLTAYSMKRMADKALVRLSPACETMGSAVTT 491  
QY 215 SLSDAKRAVDFTILYCOLVAVRETLVLVDLALILYKRNAEHVAASVENANRVRKELADTL 274

PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 14-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 18-OCT-1999; 99US-0159638.  
 PR 21-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 22-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 25-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 26-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 28-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 29-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match  
 Best Local Similarity 4.7%; Score 110; DB 21; Length 1073;  
 Matches 93; Conservatave 51; Mismatches 118; Indels 146; Gaps 20;

DB 9 LEIVLAITSAGHGRSD-----VNSLTKVERALKEASGSENEALALBGLK 55  
 DB 197 ILITIAVTSLALGKTGEGWLDGSIAPAVLTVVAVSDYRSLQ--FQNLNDEK 254  
 OY 56 GEIOTPRDRGQATKI-----LGSVGSALGKLGSDATKI----- 90  
 DB 255 RNIQLEVMRGCTVKISIDVYVGVPIPLRIGDQVPADVLLGSLAIDESSMGESKI 314  
 OY 91 -----ISGC-----LDIVAGIATFG-----CPVOMGIGAVAS 118  
 DB 315 VHKDQSPFLMSGCKVADGVGNMLVTGVTGINTGEMILMAISIEDGETPPLQVRLNGLAT 374  
 OY 119 FVSSISLFTGSSAKNSVAVIDRALSKHRDEAIOHAAGAKKRDFASSAFIOVMKQSN 178  
 DB 375 FI-GIVGL-----SVALLVVAL-----LVRIETGTDTGATQFIQGTISID 418  
 OY 179 LFDSS-----DISTAANVP-----VYKFSNFIGQ--LESRS--OQAAT 214  
 DB 419 IYDDCVKIFTTIAVTIVVAVPEGLPLAVTLTLAYSMRKMAADKALVRLSACETMSAT 478  
 OY 215 SLSDAKRAVDLILYCOLVNMRETLVLDLAIYKRGNAEHVASAVENANRYNKELEADTL 274  
 DB 479 ICSDKTGT-----LTLNMTVY-ET-----YAGSKMDVA--DNPSGLPKLVA--- 519  
 OY 275 DFLKLIP-----QALIGAVHP-----ISASETSKAILIYKRYFV 312  
 DB 520 -----LISEGVAQNTGNIGFHPDGDGEVEISSPTEKAILISMAVYKICM 562

RESULT 15  
 AAG49830 standard; Protein: 1086 AA.  
 AAG49830;  
 18-OCT-2000 (first entry)  
 Arabidopsis thaliana protein fragment SEQ ID NO: 63079.

XX XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridization assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 OS Arabidopsis thaliana.  
 PN EPI033405-A2.  
 PD 06-SEP-2000.  
 XX XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123180.  
 PR 23-MAR-1999; 99US-0123548.  
 PR 25-MAR-1999; 99US-0125788.  
 PR 29-MAR-1999; 99US-0126264.  
 PR 01-APR-1999; 99US-0126785.  
 PR 06-APR-1999; 99US-0127462.  
 PR 08-APR-1999; 99US-0128234.  
 PR 16-APR-1999; 99US-0128714.  
 PR 19-APR-1999; 99US-0129845.  
 PR 21-APR-1999; 99US-0130077.  
 PR 23-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 04-MAY-1999; 99US-0132407.  
 PR 05-MAY-1999; 99US-0132484.  
 PR 06-MAY-1999; 99US-0132485.  
 PR 07-MAY-1999; 99US-0132486.  
 PR 11-MAY-1999; 99US-0132487.  
 PR 14-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 18-MAY-1999; 99US-0134370.  
 PR 19-MAY-1999; 99US-0134768.  
 PR 20-MAY-1999; 99US-0134941.  
 PR 21-MAY-1999; 99US-0135124.  
 PR 24-MAY-1999; 99US-0135353.  
 PR 25-MAY-1999; 99US-0135629.  
 PR 27-MAY-1999; 99US-0136021.  
 PR 28-MAY-1999; 99US-0136392.  
 PR 01-JUN-1999; 99US-0136782.  
 PR 03-JUN-1999; 99US-0137222.  
 PR 04-JUN-1999; 99US-0137528.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0137702.  
 PR 10-JUN-1999; 99US-0138094.  
 PR 14-JUN-1999; 99US-0138540.  
 PR 16-JUN-1999; 99US-0138847.  
 PR 17-JUN-1999; 99US-0139119.  
 PR 18-JUN-1999; 99US-0139452.  
 PR 18-JUN-1999; 99US-0139453.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 21-JUN-1999; 99US-0139763.  
 PR 22-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.

PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135153.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140383.  
PR 23-JUN-1999; 99US-0140384.  
PR 23-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 29-JUN-1999; 99US-0141287.  
PR 30-JUN-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145224.  
PR 23-JUL-1999; 99US-0145224.  
PR 23-JUL-1999; 99US-0145226.  
PR 26-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0147038.  
PR 03-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148341.  
PR 12-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 13-AUG-1999; 99US-0149175.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149426.  
PR 18-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0152363.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154479.  
PR 20-SEP-1999; 99US-0155119.  
PR 22-SEP-1999; 99US-0155486.  
PR 23-SEP-1999; 99US-0155659.  
PR 24-SEP-1999; 99US-0156458.  
PR 28-SEP-1999; 99US-0156596.  
PR 29-SEP-1999; 99US-0157117.  
PR 04-OCT-1999; 99US-0157752.  
PR 05-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.

PR 23-JUL-1999; 99US-01452145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 28-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147302.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148365.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155658.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159741.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.

---

PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 4.7%; Score 110; DB 21; Length 1066;  
Best Local Similarity 22.8%; Pred.No.2,3; Indels 146; Gaps 20.  
Matches 93; Conservative 51; Mismatches 118;

OY 9 LPIVLAITSAKHKRSD-----VNSLTKVETALKSGSNEALELEGK 55  
DB 190 ILIIAAVTSIALGITEGLKEGMLDGSIAPFVLIVTVTAVSDFRSIQ--FOINLDEK 247  
OY 56 GEIOTKPDRVGQATKI-----LASVGSALCKLMSGDATKI----- 90  
DB 248 RNIOLEVMHNGRTVKISITDYVVGVPIPLRIDQVPADSVLTLSGHSIAIDESSMGESKI 307  
OY 91 -----ISGC-----LDVIAGIATTFG-----GPVGMGICAVAS 118  
DB 308 VHKDCKSPFLMGCKVADGVGNMLTGVGINTEWGLMASIEDGEELPPOVRNLGAT 367  
OY 119 FVSILSLFTSSAKNSVAANVIDRLSKHRDAIORHAAGAKRPAESSAFLOVKKOON 178  
DB 368 FI-GIVGL-----SVALVTVL-----LVRYPTGTODNKGATOFIKGTTSTSD 411  
OY 179 LTDS-----DLSTIANNP-----VYKFSNFICO--LESRIIS---OGAATT 214  
DB 412 IYDDCVKIETTLAVTVVAVEBGLPLATLTLLAYSMRKMADKALVRLSACETMGSAIT 471  
OY 215 SUSDRAKVADPFLLYCQLVYRETIIVLDLAILYRKGAHEVASAVENNRVELAADTL 274  
DB 472 ICSDKTGT---LTLNMOTVV-ET-----YAOGSKMDVA---DNPSGLPKLVA--- 512  
OY 275 DFLHKLPE---QALIGAYHP-----IASETSKAIIINTKYFGV 312  
DB 513 ----LISEGVAAQNTTGNIHFPHKDGGEIVEISGPSFEKALISMAWKLM 555

RESULT 14  
AAG49831  
ID AAG49831 standard; Protein; 1073 AA.  
XX AAG49831;  
AC AAG49831;  
XX  
DT 18-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 63080.  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63080.  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
FX 25-FEB-2000; 2000EP-0301439.  
XX  
FX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.



[illegible]

PR	18-JUN-1999;	99US-0139459
PR	18-JUN-1999;	99US-0139460
PR	18-JUN-1999;	99US-0139461

PR 18-JUN-1999; 99US-0139763.

PR 22-JUN-1999; 99US-0139899.

PR	23-JUN-1999;	99US-0140354
PR	24-JUN-1999;	99US-0140695
PR	28-JUN-1999;	99US-0140823

PR 30-JUN-1999; 99US-0141287

PR	01-JUL-1999;	99US-0142154
----	--------------	--------------

PR	06-JUL-1999;	99US-0142390
----	--------------	--------------

PR 09-JUL-1999; 99US-0142920.

PR	13-JUL-1999;	99US-0143542.
PR	12-JUL-1999;	99US-0142577.

PR	15-JUL-1999;	99US-0144005.
PR	14-JUL-1999;	99US-0143024.

PR	16-JUL-1999;	99US-0144086.
----	--------------	---------------

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144632.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145088

PR 22-JUL-1999; 99US-0145087

PR 22-JUL-1999; 99US-0145192.

XX DE Protein encoded by Chlamydia trachomatis pmpe gene.  
 XX KM Chlamydia; vaccine; infection; fusion protein; antigen;  
 KM pelvix inflammatory disease; trachoma; atherosclerosis; heart disease;  
 KM acute respiratory tract infection; CapI; CT529; OMCB;  
 KM polymorphic membrane protein; pmp; thiol specific antioxidant; TSA.  
 XX OS Chlamydia trachomatis.  
 XX OS  
 XX PN W0200140474-A2.  
 XX PD 07-JUN-2001.  
 XX PF 04-DEC-2000; 2000WO-US32919.  
 XX PR 03-DEC-1999; 99US-0454684.  
 XX PR 19-APR-2000; 2000US-0556877.  
 XX PR 20-JUN-2000; 2000US-0598419.  
 XX PA (CORI-) CORIXA CORP.  
 XX PI Probst P, Bhatia A, Skelky YAM, Fling SP, Scholler J;  
 XX DR WPI; 2001-374831/39.  
 XX PT Chlamydia polypeptides and fusion proteins useful for preventing pelvic  
 PT inflammatory disease, trachoma, acute respiratory tract infections,  
 XX atherosclerosis and heart disease -  
 PS Claim 2; Page 216-218; 295pp; English.  
 CC The present sequence is provided in a specification relating to  
 CC compounds and methods for the treatment and diagnosis of chlamydial  
 CC infection. The compounds provided include polypeptides and fusion  
 CC proteins comprising immunogenic portions of Chlamydia antigens  
 CC and DNA sequences encoding such polypeptides. They are useful for  
 CC vaccinating against chlamydial infection, which causes pelvic  
 CC inflammatory disease, trachoma, acute respiratory tract infections,  
 CC atherosclerosis and heart disease.  
 CC XX  
 SQ Sequence 977 AA;  
 Query Match 4.8%; Score 112.5; DB 22; Length 977;  
 Best Local Similarity 21.3%; Pred. No. 1.2;  
 Matches 110; Conservative 61; Mismatches 184; Indels 161; Gaps 28;

QY 21 GKRSDV-----NSLTKVETALKEASGSENALEALBEGKEIOTKPDVGOATKILGSGV 76  
 DB 267 GNRGNIVFYNNRCFKNVEYASSEAS-----DGAIKVTRLDVYGNRGRIFFS-- 314  
 QY 77 SALGKLSGDAKTIISGCDIYAGIATTFGPGVGMGIGAVAFVSSILSLFTGSSAKNSV 136  
 DB 315 -----DNITKNYGAI---YAPVYTLVDNGPTFYNNIANNKG 349  
 QY 137 AAVIDALSKHDEAIOHRA-----AGAKR---DEASSAFIOY 172  
 DB 350 GAIYIDGTSNKSISA-DRHAIIFENIYVYVNTNANGTSTSANPPRNAITVASSSGEILL 408  
 QY 173 MKQOS-NLTDSD-LSTIAANVPYKFSNFIGOLESRIISOGAATISLSDAKRAVD----- 224  
 DB 409 GAGSSONLIFYPDIEYSNAGVSV-SFNKEADQFGSVFSGATVNSADFHQRNLQTRTPAP 467  
 QY 225 -----FILL--YCOLVNAHETL---LVLDL---AII--YRKGNAEYVAVASAVENANVKE 268  
 DB 468 LTLNSGFLICIEHQAQLTVRFOTGTVSLGNGAVLSCKRNGCD-----SASASATLKH 523  
 QY 269 LAADTLDEHKLIPGALGAVYHPSASETSKAIILNTY-----KIFGVDPVPRPT---G 320  
 DB 524 IGLN-----LSLTKSGAEL-----PLWVEPTNNSNNYADPAATPSLSDVKLSLIDYDG 574  
 QY 321 NRRYFTNSYNTYISCEAYMGNYMFRCGSNVRNPNIRSKMSDGFYTMENSDRKRLYI 380  
 DB 321 NRRYFTNSYNTYISCEAYMGNYMFRCGSNVRNPNIRSKMSDGFYTMENSDRKRLYI 380

DB 575 NSPYESTDL---THALSSQ-----PMLISSEASNDLOQSENIIDFSGILNV 615  
 QY 381 TKHD-OG---WGNQ-TLDEPDGDOGHMR-----FILL-RHGRYV 414  
 DB 616 PHIGWGLMTWGMWAKTODPEASSATTTDPQKANRHRHTLLTLWLPAGVSPKRRSPLI 675  
 QY 415 SSKRPNWFMWESSAGYIRSMENPNPQGHWSIT 450  
 DB 676 ANTLGMNMLATESLKN-----SALFTPSGHPFWGIT 707

RESULT 12  
 ID ABB94179  
 XX ABB94179 standard; Protein; 977 AA.  
 XX AC ABB94179;  
 XX DT 05-JUN-2002 (first entry)  
 XX DE Chlamydia protein sequence SEQ ID NO:191.  
 XX KW Chlamydial infection; Chlamydia; vaccine; detection; diagnosis;  
 KW antigen; antibacterial; immunostimulant; immune response;  
 KW Chlamydia-specific T-cell response.  
 XX OS Chlamydia sp.  
 XX PN W0200208267-A2.  
 XX PD 31-JAN-2002.  
 XX PF 20-JUL-2001; 2001WO-US23121.  
 XX PR 20-JUL-2000; 2000US-0620412.  
 XX PR 23-APR-2001; 2001US-0841132.  
 XX PA (CORI-) CORIXA CORP.  
 XX PI Fling SP, Skelky YAM, Probst P, Bhatia A;  
 XX DR WPI; 2002-179901/23.  
 XX PT Novel compositions comprising Chlamydia CapI protein and its use in the  
 PT treatment of Chlamydia infection -  
 XX XX  
 PS Example 1; Page 247-249; 537pp; English.  
 CC The present invention describes compositions comprising a Chlamydia CapI  
 CC protein and methods for the diagnosis and therapy of Chlamydia infection.  
 CC Chlamydia DNA and protein sequences from the present invention can have  
 CC antibacterial and immunostimulant activities, and can be used in  
 CC vaccines. Compounds from the present invention can be used for eliciting  
 CC an immune response, specifically stimulating a Chlamydia-specific T-cell  
 CC response or inhibiting the development of a Chlamydia infection in an  
 CC animal. Methods from the present invention can be used for detecting the  
 CC presence of Chlamydia in a patient; to stimulate and/or expand T cells  
 CC specific for a Chlamydia protein; and for treatment of a Chlamydia  
 CC infection. AB192394 to AB192709 and ABB94096 to ABB94374 represent  
 CC sequences used in the exemplification of the present invention.  
 CC XX  
 SQ Sequence 977 AA;  
 Query Match 4.8%; Score 112.5; DB 23; Length 977;  
 Best Local Similarity 21.3%; Pred. No. 1.2;  
 Matches 110; Conservative 61; Mismatches 184; Indels 161; Gaps 28;

QY 21 GKRSDV-----NSLTKVETALKEASGSENALEALBEGKEIOTKPDVGOATKILGSGV 76  
 DB 267 GNRGNIVFYNNRCFKNVEYASSEAS-----DGAIKVTRLDVYGNRGRIFFS-- 314  
 QY 77 SALGKLSGDAKTIISGCDIYAGIATTFGPGVGMGIGAVAFVSSILSLFTGSSAKNSV 136  
 DB 315 -----DNITKNYGAI---YAPVYTLVDNGPTFYNNIANNKG 349



Db 663 ANTLMGNNMLATESLKN-----SALTPSGHPFWGIT 694

RESULT 8

AA683202  
ID AAG83202 standard; Protein; 964 AA.  
AC AAG83202;

05-SEP-2001 (first entry)

Protein encoded by Chlamydia trachomatis pmpe gene.

Chlamydia; vaccine; infection; fusion protein; antigen;  
acute respiratory tract infection; CapI; CT529; OMCB;  
polymorphic membrane protein; pmp; thiol specific antioxidant; TSA.

Chlamydia trachomatis.

WO200140474-A2.

04-DEC-2000; 2000WO-US32919.

03-DEC-1999; 99US-0454684.

19-APR-2000; 2000US-0556877.

20-JUN-2000; 2000US-0598419.

(CORI-) CORIXA CORP.

Probst P, Bhatia A, Skelky YAM, Fling SP, Scholler J;

WPI: 2001-374831/39.

Chlamydia polypeptides and fusion proteins useful for preventing pelvic  
inflammatory disease, trachoma, acute respiratory tract infections,

Claim 2; Page 191-193; 295pp; English.

The present sequence is provided in a specification relating to  
infection. The compounds provided include polypeptides and fusion  
proteins comprising immunogenic portions of Chlamydia antigens  
and DNA sequences encoding such polypeptides. They are useful for  
vaccinating against chlamydial infection, which causes pelvic  
inflammatory disease, trachoma, acute respiratory tract infections,  
atherosclerosis and heart disease.

Sequence 964 AA;

Query Match Best Local Similarity 4.8%; Score 112.5; DB 22; Length 964;  
Matches 110; Conservative 61; Mismatches 184; Indels 161; Gaps 28;

21 GKRSPV-----NSILTRKVEALKEAGSNEALELEGKGEIOTPPDVGATKILGSVG 76

254 GNRGNIVYNNRCRNVETASSEAS-----DGAIVTTRLDVTGNNGRIFFS-- 301

77 SALCKLNSGATKIIISGCLDIVAGIATFGSPVGMGIGAVAFSSILSLTGSASAKSV 136

302 -----DNTTNYGAI--YAPVTLVDNCPPTFFINNINANKG 336

137 AAVIDRALSKHREDAIORHA-----AGAKR-----DEASSANIOY 172

337 GATIDGTNSKRISA--DRAHITENENIVNVANSTSTSANPPRRNATIVASSSEHLL 395

173 MKQOS-NLTDSO-LSTIANVPYKFSNFISIGLESIRISGATTTSLDRAKAVD----- 224

396 GAGSSONLIFDPIEVSNAGVSV--SPNKEADDTGTSVTSAGATVNSADDFORNLQRTTPAP 454

QY 225 -----FILL--YCOLVYMRRL--LVDL-----AIL--YKGAHVAHVAENANRYNKE 268

Db 455 LTISSNGFLCTEDHQAOLTVNRFTQTGGVSLGNGAVLSCYKNGTGD-----SASNASTTLKH 510

QY 269 LAADTLDFLHKILIPBQALGANVYHPIASASETSKAILNYT-----KFGVDVPRPI-----G 320

Db 511 IGLN-----LSTLKSQAEI-----PLWVEPTNNNNNTADTAATFSLSDVLTSLIDYG 561

QY 321 NRRKFTNSYNTVYSICSEAVYNGNMFNGSCSVNRPNIRVSKMSDGFYTMESDRKLYI 380

Db 562 NSPYESTDL-----THALSSQ-----PMLTISEADNDOLQSENIDFSLG 602

QY 381 TKHD-QG--WGCG-TLDEDPDGDGHR-----FIPL-RHGRYV 414

Db 603 PHIGWGLMTWGMARKTQDDPPASATITTDOKNRFRHTLLTWLPAGYVPSRHRNSPLI 662

QY 415 SSRKMPNWMYXMESSASAGYIRSWENNPDPGHSIT 450

Db 663 ANTLMGNNMLATESLKN-----SALTPSGHPFWGIT 694

RESULT 9

ABB94173  
ID ABB94173 standard; Protein; 964 AA.

ABB94173;

05-JUN-2002 (first entry)

Chlamydia protein sequence SEQ ID NO:177.

Chlamydia; infection; Chlamydia; vaccine; detection; diagnosis;  
antigen; antibacterial; immunostimulant; immune response;  
Chlamydia-specific T-cell response.

Chlamydia sp.

WO200208267-A2.

31-JAN-2002.

20-JUL-2001; 2001WO-US23121.

20-JUL-2000; 2000US-0620412.

23-APR-2001; 2001US-0841132.

(CORI-) CORIXA CORP.

Fling SP, Skelky YAM, Probst P, Bhatia A;

WPI: 2002-179901/23.

Novel compositions comprising Chlamydia CapI protein and its use in the  
treatment of Chlamydia infection.

Example 1; Page 223-225; 537pp; English.

The present invention describes compositions comprising a Chlamydia CapI  
protein and methods for the diagnosis and therapy of Chlamydia infection.  
Chlamydia DNA and protein sequences from the present invention can have  
antibacterial and immunostimulant activities, and can be used in  
vaccines. Compounds from the present invention can be used for eliciting  
an immune response, specifically stimulating a Chlamydia-specific T-cell  
response or inhibiting the development of a Chlamydia infection in an  
animal. Methods from the present invention can be used for detecting the  
presence of Chlamydia in a patient; to stimulate and/or expand T cells  
specific for a Chlamydia protein; and for treatment and/or expand T cells  
infection. AB192394 to AB192709 and ABB94096 to ABB94374 represent  
sequences used in the exemplification of the present invention.

Sequence 964 AA;

XX The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes' their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 6281 AA:  
 SO  
 Query Match 5.1%; Score 118.5; DB 22; Length 6281;  
 Best Local Similarity 21.2%; Pred. No. 5.8;  
 Matches 64; Conservative 58; Mismatches 121; Indels 59; Gaps 12;  
 OY 16 TSARKKRSVSLITKVTALKEASGNE--AALE-----ALEGKGEIOTKPRV 65  
 DB 1808 TADREKQAYNDAYTAATLLDKTAGSNDKNAAYEQALQRYNTATYALNGD-----ERU 1862  
 OY 66 GQATKILGVSALGKLSNGDATKIISGCLDIYAGIATFEGFVGIGAVASFVSSILS 125  
 DB 1863 NNAKTAKQOVATMSHTDPAOKANLTS---QIESG--TVVAGVQGI----- 1903  
 OY 126 LTTGSSAKRSVAVIDRALSKHDEAIORHAAGAKRDFEASSAFIOWMQOOSNITPDSL 185  
 DB 1904 -----OANACTDQANMQLKOSTASKDATSSPDYQDANADLOAVAYDA--VTNAEGT 1954  
 OY 186 IIAANPYVKFSNFISGLESISOGAATTSLSDAKRAVDITLLYCOLVYVRETLVDLAI 245  
 DB 1955 ISATNPENM--PTINGKASQVN--SAKSALNGDEK-----LAAAKQTAKSIDGR 2001  
 OY 246 LYKGAHEVASAVENANRNYKELADTDFLHKLIPEQALIGAVYPIASSETSAILN 305  
 DB 2002 LTLNNQRTAANAEEVOQAPN--LAAVTA--AKNKATSLNTRAMGMLKRALAKDKNTKTSVN 2058  
 OY 306 YT 307  
 DB 2059 YT 2060  
 RESULT 7  
 AABI3634  
 ID AABI3634 standard; Protein: 964 AA.  
 AC AABI3634;  
 XX  
 DT 02-FEB-2001 (first entry)  
 DE  
 XX C. trachomatis pmpE gene protein.  
 DE  
 XX Chlamydial infection; sexually transmitted disease;  
 KW pelvic inflammatory disease; PID; tubal obstruction; infertility;  
 KW trachoma; blindness; acute respiratory tract infection;  
 KW atherosclerosis; coronary heart disease; antibacterial.  
 XX Chlamydia trachomatis.  
 OS  
 PN WO200034483-A2.  
 XX  
 PD 15-JUN-2000.

PF 08-DEC-1999; 99WO-US23012.  
 XX 08-DEC-1998; 98US-0208277.  
 PR 08-APR-1999; 99US-0288594.  
 PR 01-OCT-1999; 99US-0410568.  
 PR 22-OCT-1999; 99US-0426571.  
 XX  
 PR (CORI-) CORIXA CORP.  
 XX  
 PI Probat P, Bhatia A, Skelky YAW, Fling SP, Jen S, Stromberg EJ;  
 DR WPI; 2000-431303/37.  
 XX  
 PT Isolated polypeptide for diagnosis and treatment of Chlamydia infection  
 PT comprises immunogenic portion of Chlamydia antigen, which comprises  
 PT amino acid sequence encoded by polynucleotide sequence  
 PS  
 XX Claim 2; Pages 184-186; 256pp; English.  
 CC The present invention relates to new nucleic acid sequences and the  
 CC proteins encoded by the nucleic acid sequences. The encoded proteins  
 CC comprise an immunogenic portion of a Chlamydia antigen. The encoded  
 CC proteins are useful for the serodiagnosis and treatment of Chlamydia  
 CC infection. Chlamydiae are intracellular bacterial pathogens that are  
 CC responsible for a wide variety of human infections. C. trachomatis  
 CC infection is one of the most common sexually transmitted diseases and can  
 CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction  
 CC and infertility. Trachoma due to ocular infection with C. pneumoniae is a  
 CC the leading cause of preventable blindness worldwide. C. pneumoniae is a  
 CC major cause of acute respiratory tract infections in humans and is also  
 CC thought to play a role in the pathogenesis of atherosclerosis and  
 CC coronary heart disease. The present sequence is a protein isolated in the  
 CC present invention.  
 CC  
 XX Sequence 964 AA:  
 SO  
 Query Match 4.8%; Score 112.5; DB 21; Length 964;  
 Best Local Similarity 21.3%; Pred. No. 1.2; Indels 161; Gaps 28;  
 Matches 110; Conservative 61; Mismatches 184;  
 OY 21 GKRSV---NSLITKVTALKEASGNEALELEGLKEIOTKPRVQATKILGVS 76  
 DB 254 GNGNINIFYNNRFKNVETASSEAS--DGAIKVTRLDVTCNRRGIFFS-- 301  
 OY 77 SALGKLSGATKIISGCLDIYAGIATFEGFVGIGAVASFVSSILSLEFGSSAKNSV 136  
 DB 302 -----DNITKNYGAI--YAPVTLVDNGPTFYFINNIANKG 336  
 OY 137 AAVIDRALSKHDEAIORHA-----AGAR-----DFAESSAFIOW 172  
 DB 337 GAIYIDGTSNKSISA--DRHAIIFENITVNTNANGTSTSNPPRNAITVASSSGEILL 395  
 OY 173 MKQOS--NLTSD--LSITANPYVKFSNFISGLESISOGAATTSLSDAKRAVD----- 224  
 DB 396 GAGSSONLIFYDPIEVSNAVSV--SFNKADQTSVVFSGATVNSADPHQNLQTKTPAP 454  
 OY 225 -----FTLL--YCOLVYVRETL--LVDL--AIL--YKRGNAHEVASAVENANRYK 268  
 DB 455 LTLNNGFLICEDHADLYVNRFTQGVVSLGNGAVLSCYKNGTGD---SASNASITLKH 510  
 OY 269 LAADTDLFLHKLIPEQALIGAVYPIASSETSAILNT-----KFGVPPVPRP-----G 320  
 DB 511 IGLN-----LSITLKSQAEI-----PLTWVEPTNNSNNYTADTAATFSLSDKVLSLDY 561  
 OY 321 NRRYFTNSYNTYISCSAAYKGNMFRGCSVVRPNRFSKSGDGYTMENSDDRRLTYI 380  
 DB 562 NSPYESTDL--THALSSQ-----PMLSISASDNQLOSEINIDSGLVN 602  
 OY 381 TKHD--QG--TLEDPGDQGHM-----FIPL--RAGKIVY 414  
 DB 603 PHIGWGLMTWGAKTQDEPPASSATITDPOKANRFRHTLLTLWLPAGVYSPKRHSPLI 662  
 OY 415 SSKRWPNWFMYESSASGYIRSWENNPFGQGHWSIT 450

```

DB 1176 ISATNPEMN-PDITNOKASOVN--SAKSALNGDEK-----LAAAKQAKSDIGR 1222
OY 246 LYRKGNAEHVAVENANRVNKELAADTLDLFLHKLIPQALIGAVYHPISASSETSALIN 305
DB 1223 LFDLNNAGQTSATAEVDAPN--LAAVTA-AKKNKATSLNTAMGNLKHALLAEKDNTRRSVN 1279
OY 306 YT 307
DB 1280 YT 1281

RESULT 5
ID AU37017 standard; Protein; 5795 AA.
AC AU37017;
XX 14-FEB-2002 (first entry)
DE Staphylococcus aureus cellular proliferation protein #1187.
XX
XX Antisense; prokaryotic cellular proliferation protein;
XX antibiotic; antibacterial; drug design.
XX
XX Staphylococcus aureus.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX 23-MAY-2000; 2000US-206848P.
XX 26-MAY-2000; 2000US-207727P.
XX 23-OCT-2000; 2000US-242578P.
XX 27-NOV-2000; 2000US-253625P.
XX 22-DEC-2000; 2000US-257931P.
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX
XX WPI: 2001-611495/70.
XX N-PSDB: AAS54876.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 12610; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence represents an
XX essential prokaryotic cellular proliferation protein.
XX
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.

```

```

XX
XX SQ Sequence 5795 AA;
XX
XX Query Match 5.1%; Score 118.5; DB 22; Length 5795;
XX Best Local Similarity 20.9%; Pred. No. 5.1;
XX Matches 63; Conservative 59; Mismatches 121; Indels 59; Gaps 12;
XX
OY 16 TSARKGRSDVNSLTQVETALKEAGSNE--ALEGKGEIQKPPDRV 65
DB 4486 TPADREKOTAYNDAYTAKTILDKTAGSNDKAAVEQALORVNTAKTALNGDA-----RL 4540
OY 66 GOATKILGSVSGALGKLSNGDATKIIISGCLDIAGIATTFGGPVGMIGAVAFVSIIIS 125
DB 4541 NEAKNTAKQOQLATMSHLINAKANLTS--QIERG--TTVAVGQI----- 4581
OY 126 LFTGSSAKNSVAAVTDRLSKHRDPAIQRHAAGAKRDPRESSAFIQVKQOSNLTDSDLS 185
DB 4582 -----QANAGTLDQAMNOLROSIAKSDATKSSSEYODANADQLNNAYNHA-VSDAEGT 4632
OY 186 IIAANVPYKFSNFIQLESRISSGATTSLSDAKRAVDILLYQQLVVMRETLVLDLAI 245
DB 4633 ISATNPEMN-PDITNOKASOVN--SAKSALNGDEK-----LAAAKQAKTDIGH 4679
OY 246 LYRKGNAEHVAVENANRVNKELAADTLDLFLHKLIPQALIGAVYHPISASSETSALIN 305
DB 4680 LSDLNNAGQTSATAEVDAPN--LAAVTA-AKKNKATSLNTAMGNLKHALLAEKDNTRRSVN 4736
OY 306 YT 307
DB 4737 YT 4738

```

```

RESULT 6
ID AU37403 standard; Protein; 6281 AA.
AC AU37403;
XX 14-FEB-2002 (first entry)
DE Staphylococcus aureus cellular proliferation protein #1573.
XX
XX Antisense; prokaryotic cellular proliferation protein;
XX antibiotic; antibacterial; drug design.
XX
XX Staphylococcus aureus.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX 23-MAY-2000; 2000US-206848P.
XX 26-MAY-2000; 2000US-207727P.
XX 23-OCT-2000; 2000US-242578P.
XX 27-NOV-2000; 2000US-253625P.
XX 22-DEC-2000; 2000US-257931P.
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX
XX WPI: 2001-611495/70.
XX N-PSDB: AAS55262.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 12996; 511pp; English.

```

XX OS Chiroptalmus quadrigatus.  
XX PN JP2002125676-A.  
XX PD 08-MAY-2002.  
XX PF 24-OCT-2000; 2000JP-0323506.  
XX PR 24-OCT-2000; 2000JP-0323506.  
XX PA (SUNR) SUNTORY LTD.  
XX DR WPI: 2002-483935/52.  
XX DR N-PSDB: AAL47248.  
XX PT New hemolytically active protein useful as intermediates for  
XX PT agricultural chemicals  
XX PS Claim 2; Page 8-9; 13pp; Japanese.  
XX CC The present invention relates to proteins from the sea wasp which have  
XX CC haemolytic activity. The proteins can be used for the preparation of  
XX CC reagents, agricultural chemicals and drugs. The present sequence is a  
XX CC polypeptide of the invention.  
XX SQ Sequence 462 AA;  
Query Match 18.2%; Score 423.5; DB 23; Length 462;  
Best Local Similarity 26.3%; Pred. No. 1.9e-28;  
Matches 120; Conservative 108; Mismatches 171; Indels 57; Gaps 17;  
OY 24 SDVNSLTLYETALKEASGSEALALEGLKEIQTQKPRVQATKILGSVGS---ALG 80  
DB 26 SGLASLAKIDA--KRPSSG--KOLFDRYANMOKOIEEFSSNDRRAKMGAGISSTAVG 81  
OY 81 KLNSGDATKIISGCLDIAGIATTFGGPYGMCIGAVASFSVSIISLFTGSAAKNSVAVI 140  
DB 82 KPOSGDPAKTASGCLDILVIGISVVKD--FAKSPISITLSMVGLSGTAEESVGSV 139  
OY 141 DRALSKHRDEAIORHAAGAKRDEAFSSAFID-VAKQOONLTDLSLITAAVPPYKFSNF 199  
DB 140 KKVVOESQDDELQALYGVKREYAVSKAFIDGVNENSDISPTVSLGAVNPPYQGVRF 199  
OY 200 IGLAESISGAATTSISDAKRAVDILYQVVMRETLVLD-ALTYRKGNAEHVASA 258  
DB 200 IAAVVOIRIKNNKPRTE-SEIKRVISMELFTDLCSLRDLILDLVQLVAFPGHSPNIA 258  
OY 259 VENANRVNKEALADTLDFLKLHPEQALIGAVY-HPISASETSKAILNYTYFGVPDVR 317  
DB 259 IKVNSNGREYKRVFEDLKTNDKETYLPISLYPRRNOSOKIF---KFF---DLMK 312  
OY 318 PIGNRRYK-----FTNSYWNYSISCEAYMGNMFRGCSNVNRPNIYVSKMSDGEYT 369  
DB 313 VKYDRLKODLTGIGVSSLSHMPNVPFLCSSK---DYLLALICTKPYG-SLRDLKNDGRYS 368  
OY 370 MENDRR-----KLYITHDQGWGWTDEDPDGGHRRFPLRKG---YVSSKR 418  
DB 369 IKTKQSNPKYCHRYEYILFTFHDRN-----DLEKEKNFVPVKIGEKRIYLLSSKA 418  
OY 419 WPNMFMTESASGYIRSMENN-----GPOGHMSI 449  
DB 419 SPNKFAYVPTAKGDLFFVVDGIPSOGLGNGGYFTL 454  
RESULT 4  
AAU34339  
ID AAU34339 standard; Protein; 2434 AA.  
XX AC AAU34339;  
XX DT 14-FEB-2002 (first entry)  
XX

DE Staphylococcus aureus cellular proliferation protein #615.  
XX KW Antisense; prokaryotic cellular proliferation protein;  
XX KW antibiotic; antibacterial; drug design.  
XX OS Staphylococcus aureus.  
XX PN WO200170955-A2.  
XX PD 27-SEP-2001.  
XX PF 21-MAR-2001; 2001WO-US09180.  
XX PR 21-MAR-2000; 2000US-191078P.  
XX PR 23-MAY-2000; 2000US-206848P.  
XX PR 26-MAY-2000; 2000US-207727P.  
XX PR 23-OCT-2000; 2000US-242578P.  
XX PR 27-NOV-2000; 2000US-253625P.  
XX PR 22-DEC-2000; 2000US-257931P.  
XX PR 16-FEB-2001; 2001US-269308P.  
XX PA (ELITR) ELITRA PHARM INC.  
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Traylor JD, Carr GJ;  
XX PI Yamamoto RT, Xu HH;  
XX DR WPI: 2001-611495/70.  
XX DR N-PSDB: AAS52198.  
XX PT New polynucleotides for the identification and development of  
XX PT antibiotics, comprise sequences of antisense nucleic acids -  
XX PS Example 3; Seq ID No 5835; 51pp; English.  
XX CC The invention relates to antisense inhibitors of genes essential to  
XX CC prokaryotic cellular proliferation, their use in identifying the  
XX CC genes, their use in the discovery of novel antibiotics, the essential  
XX CC genes themselves and the encoded proteins. The prokaryotes used are  
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhimurium, Klebsiella  
XX CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
XX CC invention is also useful for the identification of potential new targets  
XX CC for antibiotic development. The antisense nucleic acids can also be used  
XX CC to identify proteins used in proliferation, to express these proteins,  
XX CC and to obtain antibodies capable of binding to the expressed proteins.  
XX CC The proteins can be used to screen compounds in rational drug discovery  
XX CC programmes. The antisense nucleic acid sequence is also useful to screen  
XX CC for homologous nucleic acids which are required for cell proliferation in  
XX CC a wide variety of organisms. The present sequence represents an  
XX CC essential prokaryotic cellular proliferation protein.  
XX CC Note: The sequence data for this patent did not form part  
XX CC of the printed specification, but was obtained in electronic  
XX CC format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pcl\_sequences.  
XX SQ Sequence 2434 AA;  
Query Match 5.1%; Score 118.5; DB 22; Length 2434;  
Best Local Similarity 21.2%; Pred. No. 1.4;  
Matches 64; Conservative 58; Mismatches 121; Indels 59; Gaps 12;  
OY 16 TSAHGRSDVNSLTLYETALKEASGSE--FALE-----ALEGLKEIQTQKPRV 65  
DB 1029 TDADREKOTAVNDAVIAKTLIDTASNDKRAVEBALORVNTAKTALMGD-----ERL 1083  
OY 66 GQATKILGSVGSALGKNSGDATKIISGCLDIYAGIATTFGGPYGMCIGAVASFSVSI 125  
DB 1084 NEAKNTAKQOVATMSHLTDQOKANLTS---QIESG--TTVAGVYGI----- 1124  
OY 126 LFTGSSAKNSVAAVYIDRALSKHRDEAIORHAAGAKRDEAFSSAFIDVAKQOONLTD 185  
DB 1125 -----QANAGTLDDAMNOLROSIAKDTKSSSEYQDANMDLQANAYNDA-VTNAEG 1175  
OY 186 IIAANVPYKFSNFIQGLAESRISGAATTSISDAKRAVDILYQVVMRETLVLDLAI 245

PT mechanism  
 XX  
 PS Claim 4; Page 26-27; 32pp; Japanese.  
 XX  
 CC This invention describes a novel: protein which has hemolytic activity,  
 CC blood platelet agglutination activity and a molecular weight of about  
 CC 50,000 Da (by SDS-PAGE). The protein, homologs and analogs of the  
 CC invention can be used as a novel approach to developing drugs useful for  
 CC the treatment of jelly fish stings, pharmaceuticals with blood platelet  
 CC agglutination activity, pesticides by use of the hemolytic activity, and  
 CC in the study of the hemolytic mechanism. This sequence represents the  
 CC hemolytic protein described in the invention  
 XX  
 SQ Sequence 450 AA:  
 Query Match 100.0%; Score 2322; DB 20; Length 450;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-195;  
 Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MILKHPMLFVLAITSAKHGRSDVNSLITKVEETALKASGSENEALFALGKGEIOT 60  
 DB 1 MILKHPMLFVLAITSAKHGRSDVNSLITKVEETALKASGSENEALFALGKGEIOT 60  
 QY 61 KPDVGOATKILGVSAGLKGNSGDATKIIISGCLDIYAGIATTTGGPGVGMIGAVASPV 120  
 DB 61 KPDVGOATKILGVSAGLKGNSGDATKIIISGCLDIYAGIATTTGGPGVGMIGAVASPV 120  
 QY 121 SSILSFTSSAKNSVAVIDRALSKHDEAIORHAAGKRDFAESSAFIOVMKQOQSNLT 180  
 DB 121 SSILSFTSSAKNSVAVIDRALSKHDEAIORHAAGKRDFAESSAFIOVMKQOQSNLT 180  
 QY 181 DSDLSITIANVYKFSNITGLESRIISOGAATISLSDAKRAVDFTLLCOLVYMERLL 240  
 DB 181 DSDLSITIANVYKFSNITGLESRIISOGAATISLSDAKRAVDFTLLCOLVYMERLL 240  
 QY 241 VDLATIKRGNAEHVASAVENANRVKELADTDLHLKLIPEQALIGAVYHPISASETS 300  
 DB 241 VDLATIKRGNAEHVASAVENANRVKELADTDLHLKLIPEQALIGAVYHPISASETS 300  
 QY 301 KAILNTKTFGVDPVPRPIGNRRYKFTNSYNTYSICSEAYMGNYMFRGCSNVNRPNIY 360  
 DB 301 KAILNTKTFGVDPVPRPIGNRRYKFTNSYNTYSICSEAYMGNYMFRGCSNVNRPNIY 360  
 QY 361 SKMSDFEYTMENSDRRKLLITTKHDGWMGTDLDEPDGOGHMFIPLRHGKYVSSKRRNP 420  
 DB 361 SKMSDFEYTMENSDRRKLLITTKHDGWMGTDLDEPDGOGHMFIPLRHGKYVSSKRRNP 420  
 QY 421 NMFMYESSASGYIRSMENNPQOGHMSIT 450  
 DB 421 NMFMYESSASGYIRSMENNPQOGHMSIT 450  
 RESULT 2  
 AAG65633  
 ID AAG65633 standard; protein; 463 AA.  
 XX  
 AC AAG65633;  
 XX  
 OS Carybdea alata.  
 XX  
 PN MO200170799-A1.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-JP02209.  
 XX  
 PR 21-MAR-2000; 2000JP-0078967.

XX (SUNR ) SUNTORY LTD.  
 PA Nagai H, Kuroda K, Nakajima T;  
 XX  
 PI WPI: 2001-611481/70.  
 XX N-PSDB: AA166621.  
 DR  
 CC New polypeptide for treating stings, and for the development of  
 PT pharmaceuticals and agrochemicals, and studying hemolysins, comprises  
 PT the Carybdea alata-originated protein -  
 XX  
 CC Claim 4; Page 22-23; 28pp; Japanese.  
 XX  
 SQ Sequence 463 AA:  
 Query Match 42.7%; Score 992.5; DB 22; Length 463;  
 Best Local Similarity 43.5%; Pred. No. 1.8e-78;  
 Matches 197; Conservative 95; Mismatches 148; Indels 13; Gaps 7;  
 QY 9 LEIVLAITSAKHG-----KRSVNSLITKVEETALKAS--GSENEAL-FALEGLKGEIOT 60  
 DB 11 LEIVLSTAFPSQARLSRYRSAAADAVSTDIDGIIQOLNDLGTFRKLEALOGVOEAVK 70  
 QY 61 KP-DVGOATKILGVSAGLKGNSGDATKIIISGCLDIYAGIATTTGGPGVGMIGAVASPV 119  
 DB 71 EATTTAKSTIVGSGSLSKFKSGDPDVASGCLDIYAGIATTTGGPGVGMIGAVASPV 130  
 QY 120 VSSILSFTSSAKNSVAVIDRALSKHDEAIORHAAGKRDFAESSAFIOVMKQOQSNLT 179  
 DB 131 ISSILSFTSSAKNSVAVIDRALSKHDEAIORHAAGKRDFAESSAFIOVMKQOQSNLT 190  
 QY 180 TSDLSITIANVYKFSNITGLESRIISOGAATISLSDAKRAVDFTLLCOLVYMERLL 239  
 DB 191 TEVHDSVDAVVAFTMMLGVLESIRINRGVSVDNNAMRTINFILYLDLSVNERLL 250  
 QY 240 LVDLALTYKR--GNEHVASAVENANRVKELADTDLHLKLIPEQALIGAVYHPISAS 297  
 DB 251 LQVILILYRAGAYDELALSLISDQKKEATREIVTLHOMETKYSICGSYIYIDHS 310  
 QY 298 ETSKALNTKTFGVDPVPR-PIGNRRYKFTNSYNTYSICSEAYMGNYMFRGCSNVNRP 356  
 DB 311 KALIGILKLTKEFPVDPARPTFDGLYRMQNRAMNRSICKESYAGNHMFRGCSNVNRP 370  
 QY 357 NIRVSKMSDGFYTMENSDRRKLLITTKHDGWMGTDLDEPDGOGHMFIPLRHGKYVSS 416  
 DB 371 GIRIKKLKENGYHTI-TLRSKAMVYTKHAOGWMGTDEDPDGOYFTFPLNGLYMTST 429  
 QY 417 KRPMPNMFMYESSASGYIRSMENNPQOGHMSI 449  
 DB 430 KRPMPDYFVYMESSAHGYIRSMHYNPDPQOMKI 462  
 RESULT 3  
 AA017909  
 ID AA017909 standard; peptide; 462 AA.  
 XX  
 AC AA017909;  
 XX  
 OS 22-AUG-2002 (first entry)  
 XX  
 DT Sea wasp hemolytic protein SEQ ID NO: 3.  
 DE Sea wasp; haemolytic; agricultural chemical; drug.  
 XX  
 KW



GenCore version 5.1.4\_P5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: May 19, 2003, 11:14:48 ; Search time 77 Seconds  
(without alignments) 778.737 Million cell updates/sec

Title: US-09-647-522-5

Perfect score: 2322  
Sequence: 1 MILKHPMLRIVLAITSARK.....SGYRSMENNPGRGHSIT 450

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

1: A\_Geneseq\_101002.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2322	100.0	450	AA133651	C. rastoni hemolytic
2	992.5	42.7	463	AA65633	C. alata hemolytic
3	423.5	18.2	462	AA017909	Sea wasp hemolytic
4	118.5	5.1	2434	AAU34339	Staphylococcus aur
5	118.5	5.1	5795	AAU37017	Staphylococcus aur
6	118.5	5.1	6281	AAU37403	Staphylococcus aur
7	112.5	4.8	964	AA133634	C. trachomatis pmp
8	112.5	4.8	964	AA683202	Protein encoded by
9	112.5	4.8	964	AB894173	Chlamydia protein
10	112.5	4.8	977	AA133640	C. trachomatis pmp

11	112.5	4.8	977	22	AA683208
12	112.5	4.8	977	23	AB894179
13	110	4.7	1066	21	AA649832
14	110	4.7	1073	21	AA649831
15	110	4.7	1086	21	AA649830
16	106	4.6	1978	20	AA727230
17	105.5	4.5	765	23	ABP25599
18	105.5	4.5	956	23	AAE33466
19	105.5	4.5	965	23	AAE33453
20	105.5	4.5	1844	18	AAW18302
21	105.5	4.5	1844	19	AAW56558
22	105.5	4.5	2504	18	AAW17871
23	105.5	4.5	2504	19	AAW56557
24	105.5	4.5	2504	22	AAW2610
25	105.5	4.5	2505	22	AAW2612
26	104.5	4.5	837	22	AAU34387
27	104.5	4.5	875	22	AAU37487
28	104.5	4.5	989	20	AAU37242
29	104.5	4.5	1788	22	AB862995
30	104.5	4.5	2086	22	AAU34143
31	103.5	4.5	349	22	AAW49636
32	103.5	4.5	565	22	AAW49642
33	101	4.3	1250	22	ABW12254
34	100.5	4.3	847	22	ABW40081
35	100.5	4.3	1645	18	AAW13502
36	100	4.3	392	20	AAW5448
37	100	4.3	927	23	ABW49720
38	100	4.3	3158	22	AAU37018
39	99	4.3	885	21	AAW32050
40	99	4.3	931	21	AAW32049
41	99	4.3	956	21	AAW32048
42	97.5	4.2	1013	19	AAW56013
43	97.5	4.2	1149	21	AAW16683
44	97.5	4.2	1445	12	AAW12108
45	97.5	4.2	2181	22	ABG05687

## ALIGNMENTS

RESULT 1  
AA133651 standard; Protein: 450 AA.

AC AA133651;

DT 06-JAN-2000 (first entry)

DE C. rastoni hemolytic protein.

KW Hemolytic protein; blood platelet agglutination; drug development;  
RW treatment; sting; jellyfish; pharmaceutical; pesticide.

OS Carybdea rastoni.

PN W09950294-A1.

PD 07-OCT-1999.

PF 30-MAR-1999; 99WO-JP01607.

PR 01-APR-1998; 98JP-0088569.

PA (SUNR) SUNTORY LTD.

PI Nagai H, Nakajima T;

DR WPI: 1999-580740/49.

XX N-PSDB: AA225610.

XX Protein with hemolytic activity, useful for drugs treating jelly fish  
XX stings, pharmaceuticals with blood platelet agglutination activity,  
XX pesticides by use of the hemolytic activity, and study of the hemolytic